SELF-AWARENESS: ISSUES IN CLASSICAL INDIAN AND CONTEMPORARY WESTERN PHILOSOPHY

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE UNIVERSITY OF HAWAI'I IN PARTIAL FILFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

IN

PHILOSOPHY

MAY 2004

By
Matthew D. MacKenzie

Dissertation Committee:

Arindam Chakrabarti, Chairperson
Eliot Deutsch
James E. Tiles
Ron Bontekoe
Catherine Sophian

ABSTRACT

In this dissertation I critically engage and draw insights from classical Indian, Anglo-American, phenomenological, and cognitive scientific approaches to the topic of self-awareness. In particular, I argue that in both the Western and the Indian tradition a common and influential view of self-awareness — that self-awareness is the product of an act of introspection in which consciousness takes itself (or the self) as an object — distorts our understanding of both self-awareness and consciousness as such. In contrast, I argue for the existence and primacy of *pre-reflective self-awareness* — a form of self-awareness that is an effect of both our embodiment and the basic structure of consciousness. In arguing for this account of self-awareness, I take up, among other things, the following: the semantics of the first-person and of indexicals in general, qualia and phenomenal consciousness; the possibility of non-conceptual self-awareness, the nature of introspection; the importance of embodiment and agency for our understanding of self-awareness, and the consequences of my account for the metaphysics of personal identity.

TABLE OF CONTENTS

PREFACE

As conscious subjects, we have the capacity to have conscious states that are *intentional*—that is, conscious states that are *about* some object. Thus I can be aware of the noise made by the car passing by my window, the diffused quality of light on a hazy afternoon, the person approaching from a distance, or that Toronto is north of Tampa. Indeed the intentionality of consciousness is so fundamental to our experience as conscious subjects that some philosophers, in both the Indian and the western traditions, have argued that intentionality is *the* defining feature of consciousness—in Brentano's famous formulation, intentionality is the "mark of the mental."

Yet, not only am I capable of being aware of objects outside myself, I am also capable of being aware of my position in relation to those objects. I can be aware of my body and aware of it *as mine*. I can be aware of certain features of my own mental states or experiences, such as their phenomenal character—i.e., 'what it is like' for me to taste a peach, feel a sharp pain, or watch a sunrise. Further, I can be aware that a passing car has been heard *by me*, and that *I* am annoyed, or happy, or in pain. Finally, I can contemplate the kind of entity I am, or hold myself responsible, or fear my inevitable demise. In short, I am capable of *self-awareness*. And what's more, self-awareness too seems to be a fundamental feature of our experience as conscious subjects.

To point out the phenomenon of self-awareness is not yet to make a philosophical point, however, but rather to raise a series of questions. If both intentionality and self-awareness are fundamental features of consciousness, what is their relation? Is self-awareness in fact a second-order intentional act that takes mental states or the self as its intentional object, or is self-awareness a different mode of consciousness altogether? Does the term 'self-awareness' refer to a single mode of

consciousness? If self-awareness and intentionality are distinct, is one more basic than the other? Could one exist without the other? Does self-awareness require a self? Does it require the possession of concepts or language?

Any adequate account of self-awareness will need to clarify the phenomenon under investigation and at least point toward answers to the above questions. Now the first step in clarifying the issue of self-awareness is to point out that the term has been used to designate a number of different phenomena in a number of different contexts. Hence it is important at the outset to specify the aspects of the issue of self-awareness with which I will be concerned.

The main forms of self-awareness I will discuss in this dissertation are awareness of one's own mental states (or features of those states), awareness of one's own body, and awareness of oneself as a subject. I take these to be our most basic forms of self-awareness, in contrast to more sophisticated forms of self-awareness such as those involved in self-contemplation or in psychological and social identity-formation. As is obvious, without the ability to be aware of one's own psychological or bodily states or to be aware of oneself as a subject, more sophisticated forms of self-awareness would be impossible.

As I will argue in Chapter 1, accounts of basic self-awareness in both the Indian and the western traditions fall into two broad categories: *reflectionist* theories and *reflexivist* theories. On a *reflectionist* account, basic self-awareness is the result of a second-order awareness taking a separate first-order awareness as its intentional object. In contrast, on a *reflexivist* account, at least some first-order mental states are self-manifesting or self-revealing, and so one's awareness of these states is not the product of a separate second-order mental state taking the first-order state as its object. Thus these reflexive states are capable of revealing themselves even as they reveal their primary intentional object. In the Indian tradition, such states are

compared to lamps in that they 'illuminate' themselves in the process of 'illuminating' their object.

Chapter 1 is primarily concerned with an analysis of these two approaches to self-awareness as they have been developed and defended in the Indian and western traditions. I examine the reflectionist theories of Locke, Armstrong, Rosenthal, and the Nyāya school. I then take up the reflexivist theories of the Buddhists, Prābhākara Mīmāṃsa, Advaita Vedānta, Kant[1], and Sartre. I then argue that theories of both types have serious problems and that any viable contemporary account of self-awareness will have to move beyond classical theories of either type.

Hence, in the remainder of the dissertation I articulate and defend a contemporary version of reflexivism; a version that draws from the classical forms of reflexivism, but also incorporates insights from recent work in philosophy of mind, philosophy of language, and metaphysics.

Chapter 2 takes up relevant issues in the semantics of the first-person pronoun. There are important conceptual links between first-person contents (mental and linguistic) and self-awareness, and by examining the particular semantics of first-person contents, we will gain insights into the structure of self-awareness and the constraints on any adequate theory thereof. In particular, I take up the much-discussed fact that first-person contents are immune to errors of misidentification. That is, for instance, when I utter "I have a toothache" it is not possible for me to be wrong about *who* has the toothache. In addition, I argue that while mastery of the use of the first-person pronoun is necessary for the development of certain relatively sophisticated forms of self-awareness, mastery of the first-person pronoun itself presupposes a more basic form of self-awareness. Moreover, this basic form of self-awareness involves non-criterial self-acquaintance, and, as such, limits the viability of models of

[1] On my reading, Kant's account of self-awareness combines elements of both reflexivism and reflectionism.

self-awareness based on criterial self-identification or an 'observation model' of introspection. Finally, I take up the relation between indexicals (e.g., 'I', 'this', 'now', and 'here') and the perspectival nature of perception in order to highlight the importance of this notion for a full account of self-awareness.

Having laid some groundwork in Chapter 2, Chapters 3 and 4 take up the two main components of my theory of self-awareness: phenomenal self-awareness and bodily self-awareness, respectively. *Phenomenal self-awareness* primarily concerns our prereflective, peripheral[2] awareness of our own (or aspects of our own) mental states and activities. As the term 'phenomenal self-awareness' suggests, I hold that the phenomenality of experience—i.e., the fact that there is 'something it is like' to have or undergo an experience, that experience involves a phenomenal character—is intimately connected to self-awareness. In particular, I defend what I call the *Phenomenal Self-Awareness Thesis.* According to this thesis, if a subject S has a conscious, phenomenal experience E of an object O, then S is also aware of E. In addition, I defend a version of the *Reflexivity Thesis*, according to which conscious experiences simultaneously reveal both the object of consciousness and aspects of the conscious experience itself. In defending these two theses, I construct a model of consciousness according to which the mind keeps track, not only of the external environment, but also of its own states and activities as well as the body. Further, I criticize and reject the 'inner observation' account of introspection in favor of an information-based account inspired by Sydney Shoemaker. Finally, after briefly touching on the issue of the ego and egoical structure of experience, I take up at length several key objections to my account of self-awareness.

The second component of my general account of self-awareness—bodily self-awareness—is addressed in Chapter 4. I take an analysis of the fact and the experience of embodiment to be crucial to any adequate account of self-awareness,

but it is especially crucial to an understanding of prereflective self-awareness. The fundamental fact of embodiment is crucial because our being located in the world is a condition for the possibility of having any experiential contact with the world at all; to perceive the world, we must be a part of the world, and we are part of the world because we are embodied. Thus the body provides both the location of our point of view and our point of departure in the world. Now while appreciation of the fact of embodiment shows the importance of our being one object among others in the world[3], appreciation of the experience of embodiment shows that we are aware of our bodies in a way that is fundamentally different from our awareness of any other objects in the world. We experience our bodies *as ours* and *from the inside.* Moreover, because we experience our bodies as a site of our subjectivity, awareness of the body (from the inside) is an important mode of self-awareness.

The experience of embodiment is of further importance because of its essential role in the correlation between perception and action. Action presupposes perception, but perception is not a matter of passive reception but of active exploration. Further, in order for perceptual information to be useful to embodied agents, it must be integrated with self-specifying information (such as the location of one's hands). After addressing the importance of bodily self-awareness, I argue that prereflective self-awareness involves non-conceptual content and respond to important objections to the very possibility of non-conceptual content.

The fifth and final chapter concerns the self. What is the role of the notion of a self in our understanding of self-awareness? Is self-awareness fundamentally awareness of *the self*? What are the implications of reflexivism for our understanding of the ontological status of the self? In order to address these questions, I examine whether the theory of self-awareness I have developed is compatible with a

---

[2] By 'peripheral' here I mean that some entity is not the main focus of one's attention.
[3] Or, for the dualist, being associated with an object in the world.

reductionist account of persons, an account which rejects the notion of a substantial enduring self. Through an examination of Buddhist and Parfitian versions of reductionism, I conclude that the reflexivist theory of self-awareness as I develop it is compatible with reductionism. I then turn to an investigation of our robust *sense* of self. In order to explain our sense of self without an appeal to any such actually existing entity as the self, I appeal to our sophisticated and multi-layered capacity for self-representation. That is, our sense of self arises, not from the existence of a self, but rather from our ability to construct various *models* of ourselves. I then suggest, following the work of contemporary philosophers such as Dennett and Flanagan, that we treat the self as kind of useful fiction, a psychological construct projected by our own self-models.

Finally, by way of conclusion, I want to briefly address some methodological considerations. I have two main goals in this dissertation. First, I want to contribute to comparative philosophy by clarifying classical Indian approaches to self-awareness and bringing them into dialogue with contemporary western views. Second, I want to contribute to the philosophy of mind in general by critically examining key theories of self-awareness and by sketching the outlines of a more adequate account. In pursuing these goals I have tried to exemplify a methodologically pluralistic approach to doing philosophy.

By a pluralistic approach to philosophy I mean one that integrates techniques, insights, and texts from a variety of philosophical styles, periods, or traditions in order to address live philosophical issues. Thus the approach should be contrasted with a purely comparative (or purely historical) approach to philosophy, in which the goal is simply to clarify or demonstrate similarities or differences between philosophical traditions. In the dissertation I not only clarify classical Indian theories of self-awareness in relation to contemporary western approaches, but also apply classical Indian insights to contemporary accounts of self-awareness (and vice versa).

In addition to the classical Indian tradition, the dissertation will engage thinkers and texts from a variety of western traditions, particularly phenomenological, analytic, and cognitive scientific approaches. I do this not only because thinkers from each of these areas of contemporary thought have insightful things to say about self-awareness, but also because I believe the intellectual tools they make available are essential to an adequate contemporary philosophy of mind. Contemporary philosophy of mind must include disciplined philosophical attention to human experience (phenomenology), conceptual and argumentative rigor (analytic philosophy), and awareness of empirical research (cognitive science). I have therefore employed these tools in approaching the issue of self-awareness.

CHAPTER 1
INDIAN AND WESTERN THEORIES OF SELF-AWARENESS

Philosophers in the Indian and western traditions have developed and defended a range of sophisticated accounts of self-awareness. My aim here is to examine several of these accounts and to draw out and assess the arguments for them. The theories of self-awareness developed in the two traditions under consideration fall into two broad categories: *reflectionist* theories and *reflexivist* theories. And while there are important differences in the way these theories are constructed and defended in Indian and in western philosophy, there are also important and instructive similarities. In light of these similarities, I have organized my discussion around reflectionist theories and reflexivist theories, rather than around philosophical traditions. As instances of the reflectionist theory of self-awareness, I have chosen to discuss Locke, the higher-order representation theories of Armstrong and Rosenthal, and Nyāya. As instances of reflexivism I will discuss Buddhism, Prābhākara Mīmāṃsa, Advaita Vedānta, Kant, and Sartre. Having assessed the main arguments for both theories, I will argue that neither theory is adequate, as traditionally formulated and defended.

1.1 Self-Awareness, Reflection, and Reflexivity

Let me begin by distinguishing three related theses about self-awareness (i.e., awareness of our own mental states).

> *The Self-Awareness Thesis*: if a subject is aware of an object, then the subject is also aware of being aware of that object.[1]
>
> *The Reflection Thesis*: self-awareness is the product of a second-order awareness taking a distinct, first-order awareness as its intentional object.

[1] This version of the thesis comes from Perrett (2003) . My thanks to him for allowing me to use his manuscript prior to publication.

> *The Reflexivity Thesis*: conscious states simultaneously disclose both the object of consciousness and (aspects of) the conscious state itself.

The *Self-Awareness Thesis* (SAT) asserts a necessary[2] connection between intentional awareness and basic self-awareness; awareness of objects entails (some sort of) awareness of awareness. SAT does not entail that the subject is *explicitly* aware of her awareness, nor does it entail that if a subject is in a conscious mental state, she *knows* that she is in that state.[3] Further, SAT, as I interpret it, should be distinguished from a weaker, dispositionalist interpretation. On the dispositionalist interpretation, if a subject is aware of an object, then, necessarily, it is *possible* that she is aware of being aware of that object. Call this the *Dispositional Self-Awareness Thesis* (DSAT). On DSAT, it is possible for a subject to have an awareness of which she is not occurrently aware—a possibility ruled out by SAT.[4] Rejecting both SAT and DSAT entails that a subject could have an awareness that is cognitively *inaccessible* to her.

SAT has been defended by a number of philosophers, both Indian and western. To some, it has seemed so obvious as to hardly need defense, while other philosophers—Indian and western—have found it almost as obviously mistaken. In any case, if one accepts SAT, one will need to give an account of the self-awareness involved. And it is here that our second and third theses enter the picture. According to the *Reflection Thesis* (RFT), self-awareness in this context is to be explained by a second-order awareness taking a distinct first-order awareness as its intentional object. In the Indian tradition, the reflectionist is one who claims that cognitions are *paraprakāśa*—illuminated (revealed) by another (awareness). Reflectionism is thus a *higher-order awareness* theory of self-awareness. Notice also that RFT is neutral as to the nature of the higher-order awareness; the awareness may be a belief, thought,

---

[2] Different versions of this thesis will posit different forms of necessity.
[3] Perrett (2003), p.6 .
[4] DSAT can also be stated negatively: a conscious state is never inaccessible to its subject.

perception, or some other form of representation.[5] Further, RFT is compatible with SAT, DSAT, and a rejection of either or both. Indeed, if one rejects SAT and holds that self-awareness is possible, one must accept some form of RFT.

In contrast, according to the *Reflexivity Thesis* (RXT) an awareness of an object reveals *itself* at the same time as it reveals its object—it is *svaprakāśa*, or self-illuminating.[6] Thus the self-awareness at issue in SAT is, on this view, a feature of first-order intentional awareness of objects; it is a form of *prereflective* self-awareness. There are, as we will see in detail below, multiple versions of RXT. On the strongest version of the thesis—held by the Yogācāra Buddhists—reflexive awareness is *self-enclosed*: awareness is *only* ever aware of itself. A more widely held (and more plausible) version of RXT holds that conscious mental states are both other-directed and reflexive. In addition, there are several possible accounts of the reflexive aspect of awareness. On the intentionalist version of RXT—held by Brentano, for instance—conscious mental states have at least two intentional objects: a (usually) distinct primary object, and itself as a secondary object. In contrast, one might hold, as Sartre does, that a reflexive awareness does not take itself as a full-blown intentional object. In any case, RXT is a *same-order awareness* theory of basic self-awareness. Note, however, that RXT is compatible with the existence of higher-order or reflective forms of self-awareness; the thesis simply denies that *all* forms of self-awareness are reflective.

In addition to awareness of one's own mental states and body, basic self-awareness includes awareness of oneself as a subject—or, in other words, awareness of one's own subjectivity. Self-awareness involves not simply awareness of one's own states, but awareness of those states, in some sense, *as one's own*. For, supposing

[5] The higher order awareness might even be an inference, as in Bhātta Mīmāṃsa.
[6] As we will see in Chapters 3 and 5, it is a contentious question what to whom or what the awareness reveals itself to. For instance, for the Yogācāra Buddhists, awareness reveals itself to itself. For the Prābhākaras, awareness reveals itself to the ātman.

for the moment that SAT is true, it is not enough to say that when a subject S is aware of an object O, that S is also aware that *someone or other* is aware of O, or that there is *somewhere or other* an awareness of O, and it turns out that it is S that is aware of O. Rather, it would seem that even minimal self-awareness involves some sense (implicit or explicit) that the state is one's own. In self-awareness, our experiences and our bodies are given to us in what the phenomenologists call the 'first-person mode of givenness'. As Owen Flanagan puts it, " . . . all subjective experience is self-conscious in the weak sense that there is something it is like for the subject to have that experience. This involves a sense that experience is the subject's experience, that it happens to her, occurs in her stream."[7] Of course, one who rejects SAT would deny that *all* experience is weakly self-conscious in Flanagan's sense, but one can nonetheless see what is meant by a minimal awareness of one's own subjectivity. Moreover, while it may be true that awareness of oneself as a subject is necessary for even minimal forms of self-awareness, we should not at the outset give too strong an interpretation of this requirement. For instance, it should not be taken to entail that one's experience is always accompanied by an explicit 'I think . . .' (or 'I see . . .', etc.). Nor should we presuppose (even if it turns out to be true) that one must have a full-blown concept of oneself in order to be minimally self-aware. I want to leave open the possibility that one could, " . . . describe what it is like to be an infant [or a non-human animal] as involving an inchoate sense that something is happening here, where 'here' gestures elusively to what the newborn will some day come to designate in more transparently self-referential terms."[8] Finally, in the discussion that follows, I will use the terms 'awareness' and 'experience', unless otherwise stated, to denote *conscious* states. The nature of consciousness is, of course, a highly contested matter. Hence, I will not attempt a definition of consciousness, so as not to beg any questions.

[7] Flanagan (1992), p.194 .
[8] Flanagan (1992), p.193 .

Thus, for instance, while I take consciousness to be a matter of degree—both as attributed to individuals and to mental states—philosophers such as Rosenthal would likely deny this. Therefore, in this chapter at least, I will leave open the question of degrees of consciousness.

1.2 Reflectionist Theories

On the reflectionist view, self-awareness—i.e., awareness of one's own awareness—is the product of a second-order awareness taking a distinct first-order awareness as its object. The higher-order awareness, as we have seen, could be an inner perception, a belief, a judgment, etc. The key point is that the higher-order awareness is distinct from its intentional object, the first-order mental state. Further, a reflectionist account of self-awareness is compatible with both SAT and its denial.[9] Thus, as we will see shortly, Locke holds both SAT and RFT, whereas the Nyāya-Vaiśeṣika school rejects SAT, while vigorously defending RFT.

1.2.1 Locke

According to Locke, consciousness is "the perception of what passes in a man's own mind,"[10] and "thinking consists in being conscious that one thinks." It is logically impossible, on Locke's view, for a subject to think[11] without perceiving that she does. He writes:

> If they say, the man thinks always, but is not always conscious of it; they may as well say, his body is extended, without having parts. For 'tis altogether as intelligible to say, that a body is extended without

[9] Though I think that, in insisting on their distinctness, there is a certain pressure for the reflectionist to say that the relation between first- and second-order awarenesses is contingent.
[10] Locke, II.i.19.
[11] The term 'think' here is used in a broad sense and Locke's comments should be taken to apply to occurrent mental states in general.

> parts, as that any thing *thinks without being conscious of it*, or perceiving, that it does so.[12]

Locke's defense of SAT is at the heart of his criticism of Descartes' belief that the soul is always thinking. For if the soul is always thinking, one must say that the soul thinks even in deep sleep. And if thinking entails consciousness of thinking—a view Descartes accepts as well—then one must be conscious of one's thinking even in dreamless sleep. What evidence is there for this? Locke thinks we have none. He admits the possibility that we simply forget the thoughts that occur in dreamless sleep, but finds the suggestion implausible and *ad hoc*. Moreover, one cannot argue that it is the *soul* that thinks in deep sleep, rather than the person, since this would, Locke argues, commit one to the odd view that there are *two* subjects of thinking, the soul and the man.

We can see here two arguments for SAT. First, Locke challenges the intelligibility of a denial of SAT. On his view, what distinguishes thoughts from other things is that they are sensible to us; they are, as we would say today, *subjective*. He writes, "Our being sensible of it is not necessary to anything but to our thoughts; and to them it is and to them it will always be necessary, until we can think without being conscious of it."[13] Since being sensible to a subject is necessary to thought, Locke thinks it is unintelligible that there could be thoughts of which the subject is unaware. Unfortunately, as it stands, his argument doesn't work. For, as the defender of DSAT will quickly point out, a thought can be necessarily *sensible* without being necessarily *sensed*. That is, a thought could be sensible without being sensed. It does not seem that Locke has a ready response to this point, but he does wonder what would be the point of a thought of which the subject is unaware.

Locke's second argument is that if we accept that there can be thinking of which the subject is not conscious, we run into problems with personal identity. This

[12] Locke (1975), p.II.i.19 .

is especially true for Locke's own account of personal identity, based as it is on the unity and continuity of consciousness. But his point is more general: if there are thoughts of which the subject is unaware, then these thoughts would be nothing for the subject. It then becomes unclear in what sense the thoughts *belong* to the subject. Is there then some *other* subject of the thoughts in question? For instance, if the soul thinks, but the person is unaware of it, then it seems there are two subjects involved. Or perhaps the subject of unconscious thoughts is somehow the mind, but not the person. In any case, Locke thinks that denying SAT leads to an unacceptable division in consciousness and hence in personal identity. Interestingly, similar considerations arise in contemporary discussions of the distinction between personal and subpersonal mental states. On some views, the subject of subpersonal states is not the person, but the mind or the brain.

Finally, a third argument can be gleaned from Locke's indirect realist account of perception. For if perceptual awareness involves direct awareness of ideas or representations and merely indirect awareness of external objects, it seems that perception involves an awareness of one's own inner states—i.e., one's ideas.

Locke's arguments are by no means decisive, though they are, I think, fairly typical of the considerations that motivate many defenders of SAT in both the western and the Indian tradition. But let me leave aside a full assessment of his arguments and turn to a sketch of his overall account of both consciousness and reflection. There are four main features I will focus on: the distinction between consciousness and reflection; his commitment to a reflectionist theory of self-awareness; his account of awareness of oneself as a self or subject; and his distinction between consciousness and the mind.

[13] Locke (1975), p. II.i.10 .

As we have seen, Locke holds that "[c]onsciousness is the perception of what passes in a man's own mind." He also holds that reflection is "the perception of the operations of our own minds within us."[14] Locke's characterizations of consciousness and reflection are quite similar, and as many commentators on Locke have pointed out, the exact nature of the distinction between them is by no means clear. Indeed, it seems that the distinction between consciousness and reflection is, for Locke, a matter of degree. One is always to some extent aware of one's ideas, but the degree of awareness can vary greatly; there is, according to Locke, "a great variety of degrees between earnest study, and very near minding nothing at all." And this applies also to perception of one's ideas and mental operations. Thus Locke remarks that one may let ideas "pass almost unregarded, as faint shadows, that make no Impression." Reflection, then, seems to be a matter of paying close (or closer) attention to one's ideas and mental operations. Moreover, on his view, reflection is required in order for the subject to form ideas about her ideas and mental operations. So, although some degree of attention is always allocated to one's ideas, a low degree of attention will not count as reflection and will not give rise to second-order ideas.

As I remarked above, if one accepts SAT, one must give either a reflectionist or a reflexivist account of basic self-awareness. Locke is clearly in the reflectionist camp. He insists that it is "impossible for any one to perceive, without perceiving, that he does perceive." Basic self-awareness is a form of inner *perception*, and as such Locke takes it to involve a distinction between the inner perception and its object. The act of perception produces an idea and that idea is in turn (necessarily) perceived by a distinct second-order perceptual act—what he calls a "reflex act of perception accompanying [the idea]."[15] It is important here to keep in mind the distinction between SAT and RXT. Even though, for Locke, all awareness involves self-

[14] Locke (1975), p. II.i.4 .
[15] Of course Locke's use of the term 'reflex' here does not commit him to reflexivism in my sense.

awareness, it is not the case that awareness is *self-revealing* or *self-illuminating*. First-order mental states *require* perception by the inner sense in order to be revealed to the subject. Of course the reflexivist will wonder what reveals the second-order act to the subject, but I will leave aside that question for now. The important point for now is that Locke holds both SAT and RFT, based on a perceptual model of self-awareness.

The third main feature of Locke's account of basic self-awareness is that it involves awareness of oneself as a subject in addition to awareness of one's ideas and mental operations. He asserts:

> For if I know *I feel pain*, it is evident, I have as certain a perception of my own existence, as of the existence of the pain I feel: Or if I know *I doubt*, I have as certain a perception of the existence of the thing doubting, as of that thought, which I call *doubt*. [. . .] In every act of sensation, reasoning, or thinking, we are conscious to our selves of our own being.

How are we to interpret this passage? If we take the first two italicized phrases as indicating the explicit content of the relevant mental states, then Locke seems to be making the point that when a subject has the occurrent thought 'I feel pain' or 'I doubt', she is in that case aware of her own being. The thoughts, after all, involve explicit reference to the subject herself. But then his insistence that *every* mental act involves consciousness of one's own being is a *non sequitur*. Just because explicitly self-referential thoughts involve consciousness of one's own being, it does not follow that *all* mental acts do so. On the other hand, we can take Locke to be saying that whenever one is aware of one's own pain (doubt, etc.) one has an awareness of oneself as the subject of that mental state. On this reading, one could be aware of oneself as subject without necessarily having a thought with explicit first-person content. In any case, Locke holds that consciousness of one's own being accompanies all mental acts. Thus Locke would reject a 'no-ownership' view of consciousness such as is defended by the Yogācārins, Hume, and Sartre.

Fourth, and finally, Locke, in contrast to Descartes, allows that there can be ideas in a person's mind of which the person is not presently conscious. At first blush, this seems to go against his general account of consciousness. Locke's version of SAT seems to commit him to the Cartesian view that all contents of the mind are present to consciousness—that is, that the mind is fully transparent to itself. However, Locke allows that a subject is not conscious of ideas "log'd in the memory." Thus, while he holds that a subject must be conscious of all her *occurrent* mental states, her occurrent mental states do not exhaust the contents of the mind. This view is also linked to Locke's argument that the mind exists even when the person is in dreamless sleep, and thus not conscious of anything at all.

1.2.2 Higher-Order Representation Theories

Recent analytic philosophy of mind has seen a revival of interest in consciousness and self-consciousness, and with it considerations of the three theses we have been discussing. In this section I want to discuss two influential versions of reflectionism that fall under the heading of 'higher-order representation' (HOR) theories of consciousness and self-consciousness. The first is D. M. Armstrong's neo-Lockean Higher-Order Perception (HOP) theory; the second is the Higher-Order Thought (HOT) theory defended by David Rosenthal.

Armstrong, like Locke, takes basic self-awareness to be a form of inner perception, though in cautious moments he says that inner sense is merely perception-like. Further, self-awareness involves two distinct acts of awareness—an inner-directed, higher-order awareness directed at a first-order mental state, e.g. an outer-directed perceptual awareness. Armstrong terms this form of inner perception 'introspective consciousness' and it involves not only awareness of ones own inner

mental states and activities but also a sense oneself as a 'continuing thing', the subject of these states.

Armstrong differs from Locke, however, with regards to SAT. Locke, as we have seen, whole-heartedly endorses SAT. In contrast, Armstrong's position entails the rejection of the thesis. Armstrong's theory distinguishes between several types of consciousness. First, a subject has *minimal consciousness*,

> if there is mental activity occurring in the mind, if something mental is actually happening, then that mind is not totally unconscious. It is therefore conscious. A single faint sensation is not much, but if it occurs, to that extent there is consciousness.

Second, there is *perceptual consciousness*, which is consciousness of what is currently going on in the subject's environment and her own body. A subject can be minimally conscious without being perceptually conscious. Third, *introspective consciousness* is "a perception-like awareness of current states and activities in our own mind."

Now whereas Locke identified consciousness with introspective consciousness, Armstrong merely says that it is consciousness 'in the most interesting sense'—that is, it is the type of conscious that typically matters to us and to which we typically refer when we talk about consciousness. Thus, although being consciousness in the full sense involves introspective consciousness, it is possible to be conscious without being introspectively conscious. As an example, Armstrong cites the case of the long-distance truck driver. In this case, the driver, having driven for a long period of time, 'comes to' and realizes he has been driving for some time without realizing what he has been doing. On Armstong's view, the driver was perceptually conscious, but not introspectively conscious. Thus, though it is atypical, we can be aware without being aware of our awareness. In this respect, Armstrong's theory is quite similar to the Nyāya view.

Furthermore, it is important to note that Armstrong distinguishes basic self-awareness from volitional introspection. Introspective consciousness is usually a kind 'reflex' inner perception—that is, in most cases it automatically accompanies our first-order mental states and activities. In contrast, 'introspection proper' involves active scrutinizing of our mental activities and may involve introspective awareness of our basic reflex introspective awareness.

Finally, on Armstrong's view, introspective consciousness serves an integrative function. The mind's higher-order control functions require a continuous awareness of (at least some of) its lower-order states and activities. This 'feedback' is crucial to the unity of consciousness and the sense that the lower-order states and activities belong to a single, continuous thing. (Armstrong leaves open what this thing might be, but since he is a physicalist he mentions the central nervous system, the brain, and the body as possibilities.)

In contrast to Armstrong, David Rosenthal rejects the inner perception version of the HOR theory and instead defends a higher-order *thought* theory of consciousness and self-consciousness. Before discussing the exact structure of his theory, however, we need to begin with some important distinctions emphasized by Rosenthal. First, we must distinguish *creature* consciousness from *state* consciousness. Second, we must distinguish between *transitive* and *intransitive* consciousness.

We can ascribe consciousness to at least two different types of things: creatures and mental states. We ascribe consciousness to a creature, roughly, if it is *awake* and its perceptual and cognitive faculties are operating. (This is not meant to be an *explanation* of creature consciousness, simply an identification of it.) It is hard to characterize state consciousness without begging any questions. Look at some

object in the room: your present perception of the object is conscious, rather than unconscious.

Transitive consciousness is awareness *of* something. It is most naturally thought of as a form of creature consciousness rather than state consciousness, since we say that a creature is conscious of some object, but we don't say that an awareness is aware of some object. Some mental state M, may be an *awareness of* some object O, but is odd to say that M is *aware of* O. Rather, M, in virtue of being an awareness of O, *makes* a creature C aware of O. On the other hand, both creatures and states can be intransitively conscious. I am intransitively conscious when I am awake rather than asleep and a mental state of mine is intransitively conscious when it is conscious rather than unconscious.

Rosenthal's HOT theory is primarily a theory of state consciousness. On his view a mental state is conscious when and only when a creature is transitively conscious *of* that state. In particular, a mental state's being intransitively conscious consists in the subject of that state having a higher-order thought that she has or is in that state. "The core of the theory, then," writes Rosenthal, "is that a mental state is a conscious state when, and only when, it is accompanied by a suitable higher-order thought."[16]

What is the nature of this 'suitable' higher-order thought? Rosenthal requires that a subject's consciousness of the mental state is not the product of inference or some form of external observation. This is meant to rule out cases in which one discovers that one possesses a certain mental state by being persuaded by the testimony of others or by observation of one's own behavior. Further, the HOT has assertoric content—it is an assertion that the subject has or is in a particular first-order mental state. As such, the HOT will have indexical, first-person content. Thus, for

[16] Rosenthal (1999), p. 731

instance, my pain is conscious just in case I have a HOT that 'I am in pain'. Hence, as with Locke and Armstrong, basic self-awareness involves not just awareness of one's own mental states, but also awareness of oneself as the subject of those states. In addition, the HOT is not intransitively conscious. It makes its subject conscious of a first-order mental state, but it is not itself conscious. Rosenthal must take this position in order to avoid the charge of circularity. For, if he attempts to *explain* intransitive state consciousness by an appeal to HOT and the HOT is itself conscious, his explanation would rely on what he is trying to explain.

Having sketched the two main versions of HOR, I want to conclude by pointing out three main characteristics of such accounts. First, the HOR theories we have considered rest on the primacy of transitive consciousness. That is, intransitive state consciousness can be explained in terms of transitive creature consciousness. Both Armstrong and Rosenthal intend to give a substantive, reductive account of consciousness. Hence they attempt to explain intransitive consciousness in terms of a more basic notion, in this case transitive creature consciousness. This is in contrast with Locke, who did not seem to be after a reductive account of consciousness, but was rather giving an account of what he took to be certain necessary features consciousness.

Second, HOR theories are committed to what we may call the *independence condition*: the presence of any given mental state is independent of the subject of the state being introspectively aware of that state.[17] This is, of course, a denial of SAT. According to HOR theories the intransitive consciousness of a mental state is an extrinsic property of that state. Thus it must be possible for that state to exist whether or not its subject is aware of it. Again, this is tied to the attempt to give a reductive account of consciousness. A state can be intrinsically mental, but if one is to give a

[17] Rowlands (2001),

reductive explanation of intransitive state consciousness in terms of one mental state taking another as its object, neither state can be intrinsically conscious.

Third, and finally, HOR accounts rely on the explanatory primacy of vehicles of awareness. That is, certain fundamental changes in the character of one's awareness of the environment can be explained in terms of changes in one's awareness of the properties of the vehicle of one's awareness (of the environment). This is particularly clear in Armstrong. Recall the case of the truck driver. The way the driver's experience seems to him—indeed that it seems to him to be any way at all—is explained in terms of his introspective awareness of certain features of his own perceptual states. Thus, for instance, the vividness of his experience of driving is to be explained in terms of his inner perception of his own perceptions and bodily activities.

1.2.3 Nyāya

The last version of reflectionism I will consider is that of the classical Indian school of Nyāya.[18] According to Nyāya, there is no necessary connection between awareness and awareness of awareness, and introspective awareness is explained as a form of inner perception or apperception called *anuvyavasāya. Anuvyavasāya* is a form of inner perception carried out by an inner sense organ called the *manas.* And of course, apperceptive cognitions are strictly distinct from non-apperceptive cognitions (*vyavasāya*) such as outer-directed perceptions or occurrent thoughts.

Further, the Nyāya account of apperception is not an account of consciousness *per se*, but only an account of introspective awareness. Thus, unlike HOR theories, the Naiyāyikas do not hold that consciousness consists in or is explained by apperception. For them *vyavasāyas* are conscious whether or not they are

[18] In fact the views I will discuss under the heading of Nyāya are also representative of their sister school, Vaiśeṣika.

apperceived. Furthermore, while it is not uncommon for our first-order mental states to be apperceived, on the Nyāya view, it would be too strong to say that our mental states are automatically apperceived. Thus *anuvyavasāya* is somewhat closer to Armstrong's 'introspection proper' than to his 'reflex' introspective awareness. For Nyāya, apperception involves the focussing of attention on one's own states.

Apperception is said to occur only after the arising of a first-order state, since the connection between a *vyavasāya* and an *anuvyavasāya* is causal and cause precedes effect. And given this causal-perceptual account of introspection, the Naiyāyikas argue that no mental state can be self-illuminating because that would entail the absurd consequence that the mental state was somehow self-caused. In addition, the content of an apperceptive awareness will involve indexical first-person content.[19] Thus, for example, a first-order perception of a pot would be expressed as 'there is a pot', while an apperception of that cognition would be expressed as 'I perceive a pot'. Indeed, the Naiyāyikas argue, given that self-awareness involves first-person content, but not all perception does so, it cannot be the case that all perception involves self-awareness.

It will be instructive at this point to consider some of the Naiyāyika's counter-examples to SAT. It is possible, on the Nyāya view, for a subject to have an occurrent, conscious mental state of which she is unaware. For instance, if a person sees a blue pot, but her attention is immediately shifted elsewhere, the perception of the blue pot could disappear before an apperceptive awareness of it could arise. Or, the subject might suddenly die before she could notice that she is aware of a blue pot. In both cases, there is awareness without awareness of awareness because the subject does not notice her own states. Note however, that on the Nyāya view, the subject is

[19] Actually, the later Navya-Nyāya (Neo-Nyāya) theorist, Gaṅgeśa holds that first-person contents require two apperceptive awarenesses. The first awareness apprehends the cognition in impersonal terms, i.e., 'A perception of a pot has arisen'. The second links the first-order awareness to the self and takes the form, 'I am aware of a pot'.

still conscious and is aware of the object of her perception. Next, there are cases of what we might call absorption. For instance, I might be so transfixed by an object of perception that all of my attention is locked on the object. In such cases, Nyāya holds that I am only aware of the object, and not of my own states or myself. In order to be aware of myself (including states of myself), I would need to allocate some of my attentional resources to my own states, rather than exclusively to states of the world. And, on the Nyāya account of this process, it would involve using the *manas* to perceive my own inner mental states.

The Nyāya theory, then, is committed to the independence condition— the presence of any given mental state is independent of the subject of the state being introspectively aware of that state. However, Nyāya does not rely on the primacy of transitive creature consciousness in the way that HOR accounts do. That is, the Naiyāyikas do not explain state consciousness in terms of a subject's being aware of her first-order mental states. Indeed, the Naiyāyikas do not explain state consciousness at all in their account of apperception. Rather, the consciousness of first-order mental states is a given, and what is explained is our awareness of those conscious states. In another sense, though, one can say the Nyāya theory places primacy on transitive consciousness. On the Nyāya view, the essential nature of a cognition is to reveal its object; awareness is essentially intentional. Moreover, the Naiyāyikas hold that the intentional arrow of a cognition is—with a few special exceptions—exclusively other-directed. Thus, they argue that a mental state can no more reveal itself than a knife can cut itself. Finally the Naiyāyikas would, I think, deny the explanatory primacy of vehicles. In contrast to Armstrong's view, the primary explanation for changes in how one's experience seems to one will be due to how things are in the world. This is in part because the Nyāya account of perceptual consciousness takes it to be radically transparent: perceptual awareness has no real

structure of its own, but simply reveals (when veridical) worldly objects just as they are.

1.3 Arguments for Reflectionism

Now that I have discussed a few of the main versions of the reflection theory, I want to draw together some of the most important and most influential arguments for reflectionism. I am pulling together these arguments from the (sometimes quite different) theories we have discussed so far, but one finds the same or similar arguments occurring and reoccurring across time-periods and traditions. Moreover, since not all versions of reflectionism are the same, and since, as we have seen, reflectionism is compatible with both SAT and its denial, the arguments I'll discuss do not form a consistent set. Nonetheless, consideration of these arguments allows one to see some of the key considerations in favor of reflectionism. So let me begin by listing the arguments to be considered below: phenomenological arguments, content arguments, the intentionality argument, the sensory modalities argument, the explanatory argument, the distinct existences argument, and the anti-idealism argument.

Phenomenological arguments for RFT appeal to certain aspects of our experience or to possible forms of experience in support of reflectionism. The Naiyāyikas, as we have seen, offer two key phenomenological arguments in support of their version of reflectionism. First, they point out that our experience does not always involve explicit (my term, not theirs) forms of self-awareness. There is a salient phenomenological distinction between seeing or thinking about an object in the world and being aware of or thinking about one's awareness of that object. In particular, our basic perceptual awareness seems to lack any explicit first-person content. To the extent that our experience displays the former character without

always displaying the latter, one can argue that (a certain construal of) SAT is untenable. Second, beyond our normal experience of awareness without reflection on awareness, certain types of absorption, it is argued, are exclusively other-directed and do not involve self-awareness at all. Similarly, Armstong's appeal to the long-distance truck-driver is meant to show a relatively familiar type of experience wherein awareness and awareness of awareness come apart.

Although important, these phenomenological arguments do not seem to me to be decisive. It is still open to the defender of reflexivism (or SAT)[20] to attempt to redescribe the cases or to clarify her account of self-awareness in such a way that it avoids them. For instance, if one is a Sartrean reflexivist, one will not be troubled by the Nyāya arguments because the Sartrean view does not claim that all awareness involves explicit attention to one's own states. In the case of the truck-driver, the reflexivist could construct an alternative interpretation of the case that does not support reflectionism—as I do in Chapter 3.

There are two *content arguments* in favor of reflectionism. In combination with its phenomenological argument, Nyāya points out that the proper content specifications of basic perceptions do not involve mention of the subject—i.e., the content of my basic perception of a pot will be most properly specified as 'There is a pot' rather than 'I see a pot'. The pot, not my self or my states, is the object of the perception on the first specification, whereas in the second, it is my seeing of the pot that is the object. Failure to distinguish 'there is a pot' from 'I see a pot' gives unwarranted support to reflexivism. Secondly, if the reflexivist takes the content of an awareness to be a proper part[21] of that awareness *and* we take all awareness to be (partially) about itself (as Brentano seems to), then we must hold that the awareness is a proper part of itself.

[20] Notice that these defenses of RFT are based on counter-examples to SAT. If one accepts that self-awareness is possible and one rejects SAT, then one is left with RFT by default.

The first content argument is effective against versions of reflexivism, such as the Prābhākara view, that hold that all awareness involves explicit first-person content. However, since reflexivism does not entail such a view, the argument is limited in scope. The scope of the second argument, on the other hand, may be too wide. It seems to rule out *any* form of self-referential content. Take a thought such as 'This very thought is occurring now'. A strange thought, to be sure, but is there something incoherent about it? Perhaps this argument says more about the problems with understanding the relation between content and awareness mereologically than it does about reflexivism.

The *intentionality argument* is closely related to the content argument. On a positive construal of this argument, the reflectionist can argue that her model of self-awareness dovetails nicely with the independently plausible idea that most—if not all—of our mental states are essentially intentional. And since intentionality is usually or paradigmatically taken to be directedness toward an object *distinct* from the awareness of that object, our normal conception of intentionality supports reflectionism. My awareness of my own mental states, if it is intentional, is most plausibly understood as the product of an inner-directed awareness taking a distinct awareness as its intentional object. On a negative version of the intentionality argument, the reflectionist can argue that reflexivism is incompatible with the intentionality of the mental. For instance, if one could show that *all* awareness is intentional and that intentionality is an *exclusively* other-directed relation, then reflexivism will be difficult to defend. The intentionality argument is one of the more powerful in the reflectionist arsenal, and I will discuss the issues raised by it at greater length in Chapter 3. At this point, though, two reflexivist responses might be apposite. First, with regard to the positive version of the intentionality argument, the reflexivist

[21] A proper part here is a part *not* identical to the whole.

may simply deny that we should model inner awareness on outer awareness. Second, with regard to the negative version, the reflexivist can admit that every complete awareness—she might say, every *experience*—is intentional. It is just that, on the reflexivist view, every intentional experience also involves a self-revealing aspect.

According to advocates of HOR theories, some version of their theory is the best (perhaps even the only) hope for giving a substantive explanation of consciousness and/or self-awareness. The idea behind this *explanatory argument* is that a substantive account of state consciousness cannot appeal to conscious states as part of the explanation. What can it appeal to? HOR theories appeal to the relations between non-conscious mental states in order to explain consciousness. Further, if consciousness—at least 'in the most interesting sense'—consists in a subject's being aware of her own mental states *and* we take conscious states to be self-illuminating, then we will be unable to give the kind of substantive account of consciousness offered by HOR theories. This, they argue, will likely be a net loss in the explanatory power of our account of consciousness.

Of course, it is open to the reflexivist to reply that she is not interested in giving a reductive explanation of consciousness. On the other hand, she could argue that HOR theories have other theoretical problems that outweigh their ability to give a substantive account of consciousness. Thus, the reflexivist may decide that, all things considered, she should reject HOR and look elsewhere for a substantive account of consciousness. Further, it should be noted that although HOR accounts attempt a reductive account of intransitive state consciousness, Armstrong appeals to other forms of consciousness in his account and Rosenthal relies on transitive consciousness in order to get his HOT theory off the ground. Thus one might legitimately question the explanatory robustness of HOR accounts.

The *distinct existences* argument is employed by both Nyāya and Armstrong. It is reasonable to assume that in a broad range of cases our awareness of an object is (in some sense) *caused* by that object. For example, my computer is part of the causal complex responsible for my present visual awareness of it. Further, it is plausible to assume that an apperception of some mental state M is caused by M. And, as both the Naiyāyikas and Hume have pointed out, cause and effect must be distinct existences. No event, hence no mental event, can cause itself. Thus, it is plausible to hold that second-order mental states are distinct from first-order mental states, simply because the latter are objects of the former, and objects are causes. The main problem with this argument is not that cause and effect must be distinct existences, but rather that the reflexivist is very unlikely to accept that our basic awareness of our own mental states is *caused* by those states. Thus the reflexivist can reply that the connection between conscious states and awareness of those states is not *causal*, but is rather a relation of inclusion or of identity.

Finally, with regard to the *anti-idealism argument*, it has been maintained by the Naiyāyikas that reflexivism leads to or lends support to idealism. Thus to the extent that idealism is a problematic or even incoherent philosophical position, one should reject reflexivism in favor of reflectionism. The identification of awareness with a form of self-awareness, it is argued, pushes us in the direction of (though it need not *entail*) seeing consciousness as either primarily or even exclusively self-directed. A case in point is the radical idealism of Yogācāra Buddhism. Indeed the Yogācārins use reflexivist arguments to *support* their brand of idealism. In addition, a contemporary reflectionist might point to Berkeley, Fichte, and Husserl as further evidence of a link between reflexivism and idealism. But what, exactly, is the connection here? Imagine that one holds that consciousness is essentially reflexive and one defends a representationalist account of perception. On this view perceptual

consciousness involves a direct awareness of one's own mental states and a merely indirect awareness of external objects of perception. (This is basically the view of the Sautrāntika Buddhists.) Consciousness, then, is primarily self-directed; and it is a short step from here to the idea that the external objects of perception are merely projections, or constructions out of sense-data, or what-have-you. That is, reflexivism plus a 'veil of ideas' account of perception is fertile ground for the development of idealism, as is demonstrated in both the Indian and western traditions. In contrast, the Nyāya view, with its direct realist account of perception and its insistence that *intentionality* not reflexivity is the mark of consciousness, is inhospitable to the development of idealism.

I think the Naiyāyikas have a point here. However, the reflexivist can respond that, despite their dire warnings, the Naiyāyikas do not show that reflexivism *entails* idealism. Theirs is merely the charge of giving aid and comfort to the enemy (if idealism *is* an enemy). Further, if historical examples are appropriate here, the reflexivist can point to Aristotle, the Prābhākaras and Sartre as relevant counter-examples. Reflexivism is compatible with a robustly realist account of intentionality and does not entail that reflexive awareness is completely or even primarily self-directed. Reflexivism might be one path to idealism, but reflexivists need not become idealists.

While none of these arguments seems to me conclusive, taken together they present a plausible case for a reflectionist account of self-awareness. Another (argumentative) strength of the theory is that it might be true whether or not the self-awareness thesis is true. In contrast, as we will see below, the reflexivist is committed to some version of SAT. Hence the reflexivist must not only argue for reflexivism directly, she must also defend SAT against reflectionist criticism of that thesis. It is the reflexivist theory of self-awareness to which I will now turn.

1.4 Reflexivist Theories

As we have seen, the reflexivist holds that at least some types of awareness are self-revealing or self-illuminating (*svaprakāśa*). On this view, in order for an individual to be acquainted with one of her own mental states it is not necessary that that mental state become the object of a distinct, second-order mental state. Thus, some cognitions have the power to reveal themselves, perhaps in the process of revealing their intentional object. Further, the reflexivist does not need to hold that *all* mental states are self-revealing, though some do hold this strong interpretation of RXT. In addition, some reflexivists hold that all awareness is explicitly self-referential—i.e., that it makes explicit reference either to the awareness itself or to the subject, or both—while others hold the reflexivity of awareness merely consists in a prereflective acquaintance with the awareness that does not entail explicit self-reference.

1.4.1 Buddhist Reflexivism

According to the Buddhist reflexivists—such as Dignāga, Dharmakīrti, and Śāntarakṣita—consciousness is like a lamp, revealing itself in the process of revealing its object. Their term for reflexive awareness (or self-awareness) is *svasaṃvedana*. In giving this sketch of *svasaṃvedana*, I want to focus on Dharmakīrti's formulation of the notion. And while the idea of *svasaṃvedana* is most closely associated with the Yogācāra school of Buddhist philosophy, one need not, of course, accept the Yogācārins' idealist ontology in order to accept that awareness is reflexive. Indeed, the notion of *svasaṃvedana* would also be acceptable to realists such as the Sautrāntikas. Now, in order to understand *svasaṃvedana*, we need to start with Dharmakīrti's account of perception.

Dharmakīrti defended a basically representationalist theory of perception. On his view, the object of perception leaves its 'mark' on awareness, and awareness thereby captures or reflects the form of the object. Hence, awareness takes on a certain structure or aspect (*ākāra*) due to a causal link between an object and a sense organ. This aspect constitutes what I will call the 'aspectual structure' of the awareness, the particular way the awareness is structured by and reflects its object.

Take, for instance, my visual perception of a cup. The particular aspect (*ākāra*) of this awareness might involve my seeing the object *as* a cup, *as* red, and *as* within my reach. And it is important to note here that the aspect is a feature of the *awareness*, not a feature of the object of awareness. The aspect is to be understood as the way the object *appears* to me, which may or may not reflect how the object actually is. Thus, two people could have experiences with exactly similar aspects (say, *as of* a red cup), even though one person's experience is veridical while the other's is an hallucination.

Moreover, on Dharmakīrti's view, it is only *in virtue* of an awareness having a certain aspectual structure—a certain representational character—that we have experiential access to the object at all. That is, if there were no particular way the object is represented *in* awareness, one would simply not be aware of an object in the first place. Indeed, Dharmakīrti argues that we cannot have *direct* access to objects of awareness at all, but only access mediated by an aspect. Hence, he takes perception to be the *direct* apprehension of an aspect, which then makes possible the *indirect* awareness of the external object.[22]

On this account, therefore, perception involves two components: the aspect and the apprehension of the aspect. Dharmakīrti refers to the former component as the *objective aspect* (*grāhyākāra*), and the latter as the *subjective aspect* (*grāhakākāra*).

[22] Of course, as a Yogācārin, Dharmakīrti believes that, in the final analysis, there are no external objects. But in arguing for his general account of perception, he does not, for the most part, rely on this ontological claim. Thus, the general view I am sketching here is compatible with a realist form of representationalism, such as was defended by the Sautrāntikas.

The objective aspect of an awareness constitutes the intentionality or object-directedness of that awareness. This aspect also involves the particular way in which the awareness represents its object. In contrast, the subjective aspect constitutes one's awareness of how the objective aspect represents the object.

Thus, on Dharmakīrti's view, intentional awareness fundamentally involves self-awareness (*svasaṃvedana*) in that one is aware not only of an object, but also *how* that object is given in awareness. And, of course, the subjective aspect is not a separate, second-order awareness that takes a first-order awareness as its object. Rather, the subjective aspect is a feature, an aspect, of the first-order awareness itself. Awareness reveals *itself* in the act of revealing an object. Hence, awareness is essentially reflexive and needs no separate second-order act of awareness to reveal its objective aspect. Moreover, since this type of self-awareness is not a separate awareness in its own right, it lacks the dyadic, subject-object structure of a full-blown awareness. That is, the subjective aspect of an awareness does not itself further divide into subjective and objective aspects. This is why Dharmakīrti characterizes *svasaṃvedana* as a form of non-dual awareness.

But why does Dharmakīrti hold this view? Since Dharmakīrti already holds that we are only aware of the object of an awareness by means of the aspectual structure of the awareness, he argues that self-awareness is precondition of awareness as such. He says, "If cognition were not itself perceived, perception of an object is never possible."[23] The idea here is that consciousness is, in Locke's phrase, perception of what passes in one's own mind. So if I fail to perceive what passes in my own mind, I can't be aware of external objects at all. Moreover, Buddhist reflexivists argue that reflexivity and intentionality are the distinguishing features of consciousness. Unlike insentient objects or processes, consciousness reveals itself by

[23] Quoted in Mokṣākaragupta (1985), p.51 (18.16) .

its very occurrence. Śāntarakṣita develops this point by arguing that an awareness of which one is totally unaware would not count as an awareness at all. In the complete absence of awareness of how an awareness is representing its object, the so-called awareness of the object is, according to Śāntarakṣita, no more cognitively effective than a stone.

Suppose, on the other hand, that in order to be aware of how a particular awareness represents its object a subject must have a separate, second-order awareness of that first-order awareness. (Note that these are conscious cognitions.) That is, in order to be aware of how the first awareness (A1) represents its object, there must be a second awareness (A2) that represents the representational character of A1. But then, Dharmakīrti will ask how the subject knows how A2 represents A1. If it is replied that a third awareness is needed, we are off on a vicious regress. On the other hand, if the subject doesn't know how A2 represents A1, she will not know how A1 represents its object. And if she doesn't know how A1 represents its object, in what sense, he will ask, is she aware of the object at all? Dharmakīrti's view attempts to avoid the regress problem by holding that the subject's awareness of how her awareness represents its object is not a separate, second-order intentional awareness, but is instead *part of* the first order awareness itself. The regress must be stopped somewhere by a self-intimating awareness, so why not allow that first-order awareness itself is reflexive? Consciousness, then, is self-intimating in that intentional awareness of objects and non-dual awareness of awareness are two sides of the same coin.

In addition, Dharmakīrti argues that our experience involves an affective tone (pleasant, unpleasant, or neutral) which reveals itself in the apprehension of intentional objects. And since (i) the feeling tone of an experience is a property of the experience, not the object of the experience and (ii) the feeling tone is apprehended

simultaneously with the object, it follows that the subject is aware of both the object of awareness and the awareness itself. Hence, *svasaṃvedana* involves a non-dual awareness of how the object is presented to awareness, both in the sense of how the object is represented and how it *feels* to be aware of the object.

Furthermore, Dharmakīrti argues, following Dignāga, that the phenomenon of memory gives support to a reflexivist theory of self-awareness. When I remember something—say, a sunset I saw ten years ago—I remember not just the object, but my *experience* of the object. And I remember the object and the experience *together*, since they are for me inseparable. Thus, for instance, in remembering the sunset, I can also at the same time remember how it felt to see that sunset. Therefore, the Buddhists argue, it is plausible to hold that at the time when I saw the sunset I was aware both of the object of perception and my awareness of that object.

Now, because the subjective and objective aspects of consciousness are so intimately connected, Dharmakīrti argues that the subjective aspect's apprehension of the objective aspect is infallible. As Mokṣākaragupta puts it, "[*svasaṃvedana*] is called indeterminate knowledge free from fictional constructs and unerring, because its nature consists in the direct intuition of the nature of itself."[24] Consciousness essentially involves infallible self-acquaintance. Yet this does not mean that we can never in any sense be wrong about our mental states. For example, as a Yogācārin, Dharmakīrti himself holds that we are mistaken in thinking that our mental states apprehend real, external physical objects, since, on his final view, there are no such objects. But he holds that *svasaṃvedana* is necessarily valid independent of the validity of *outer* perception because it reveals how our perceptions *seem*, and it's hard to see how we could be wrong about how a perception *seems*. The subjective aspect

[24] Mokṣākaragupta (1985), p.47 (16.2) .

of an awareness reveals the objective aspect just as it is, regardless of whether the objective aspect is itself veridical.

So, on the Buddhist reflexivist view experiences are internally complex, involving both an other-directed objective aspect, and a self-directed subjective aspect. However, in light of the view that awareness is self-revealing, it is reasonable to wonder *to whom* awareness is revealed. As we will see below, the Prābhākara reflexivists hold that a cognition reveals its intentional object and itself *to the Self* (*ātman*). However, the Buddhists deny the existence of the Self, and so cannot say that awareness reveals itself to the Self. On the Buddhist reflexivist view, consciousness reveals aspects of itself *to itself.* There is no Self or ego over and above (or, perhaps, *behind*) the stream of experience, viewing the passing show, as it were. The Self is not acquainted with its experience, on the Buddhist view; rather, experience is *self-acquainted.* The stream of consciousness (*cittasaṃtāna*) is constituted by an interconnected series of experiences, each one having a subjective aspect that is acquainted with its own object aspect. The reflexivity or self-referentiality of experience contributes to the individual's[25] sense of self, the *ahaṃkāra. Svasaṃvedana* involves an awareness of what one's experience is like both in the sense of how the experience represents its object and how it feels to undergo the experience. And this awareness typically involves a sense that the experience is the subject's own, that it happens to her or occurs within her stream of experience. *Svasaṃvedana* thus provides a continuous, immediate, and internal first-person perspective on one's own stream of experience, which reinforces the individual's sense that there is a Self or Subject that exists over and above this on-going stream.

[25] It will be objected that the Buddhist cannot deny the Self and yet appeal to the 'individual'. However, the Buddhist will give a reductionist account of the individual person (*pudgala*) while rejecting the notion of a Self or (*ātman*). We are, on this view selfless persons.

1.4.2 Prābhākara

According to the Prābhākara Mīmāṃsas, experience has a triadic structure (*tripuṭīpratyakṣatā*); a single experience reveals its primary intentional object, the experience itself, and the subject of the experience. Thus, for instance, my visual experience of a blue pot constitutively involves awareness of myself as aware of the blue pot. In addition, my desire for lunch constitutively involves awareness of this desire; the desire is self-manifesting, but unlike a cognition, it is not *about* itself, since my desire for lunch is not at the same time a desire for my desire for lunch.

One of the central motivations for Prābhākara reflexivism arises from considerations of the linguistic expression of mental states. According to the Prābhākaras, "That a cognition has productive power with regard to the linguistic usage characteristic of it is what we mean when we say that it has itself as content."[26] On this view, our first-order mental states, such as a perception of a pot, are immediately reportable. That is, if I am presently seeing a pot, I can report what I am seeing directly, without the intervention of a second-order act of introspection to reveal my first-order state to me. Thus, if a first-order state is immediately reportable, that implies that the subject of that state is acquainted with that state without the need of a second-order 'inner look'—that is, the first-order mental state is self-revealing. Further, notice that the cognition has itself as its own content only in the sense that it has the power to produce the appropriate linguistic expression of the cognition's content. The primary content of an awareness of a pot, then, is the pot, and the awareness of the pot has itself as part of its content only in secondary sense.

This distinction, however, raises the important problem of giving the correct account of the content of a first-order mental state. When I am presently aware of the pot and I am asked what I see, I might answer "A pot." However, I might also answer

[26] Phillips (2001), p.3 . Although accurate, not that this is characterization of the Prābhākara view by their Nyāya opponents.

"I see a pot." On the Prābhākara view, a normal first-order awareness of a pot has the power to produce either of the above expressions. But which one more accurately expression the actual content of the awareness? For the Prābhākaras, it is, of course, the latter expression that best captures the real content of the awareness. On their view,

> All cognitions arise in the form "I cognize this," with the cognizer, the cognition, and the cognized all as explicit. Among these, a self appears as the agent, a cognized as the object, and the cognition itself as the act. Thus the position of those who hold that perception is threefold: a cognition is experienced with a cognizer, a cognized, and itself all as content.[27]

It is because of the triadic nature of perception that a first-order perception can give rise to the expressions of the form "I perceive this." Moreover, according to the Prābhākaras, awareness of oneself and one's awareness is part of the *explicit* content of awareness. Thus their view is stronger than Kant's requirement that 'I think' should be *able* to accompany all one's representations, since the Prābhākaras hold that 'I think (cognize)' *always* accompanies our representations. On the other hand, our linguistic expression of the content of our awareness may be incomplete, only mentioning the primary intentional object of the awareness.

Finally, in addition to direct reportability, the Prābhākaras give three other important considerations in favor of their version of RXT. First, the Prābhākaras argue that their view has the benefit of parsimony. They are able to explain the reportability of a cognition without appeal to an extra layer of representation, such as the Naiyāyika's *anuvyavasāya.* Second, the Prābhākaras argue that the inner perception account of self-awareness is problematic because consciousness cannot become a perceptual *object.* Consciousness, like the ether, is imperceptible. The self-illumination view, however, does not need to appeal to this problematic form of inner perception, and so is, according to the Prābhākaras, a better view. Third, the

Prābhākaras appeal to the problem of regress. The basic idea of their version of the regress argument is this.[28] If a subject's awareness of a first-order cognition C1 requires a second-order apperception C2 of C1, then what reveals C2 to the subject? If a third cognition C3 is required, then there is a vicious regress. On the other hand, how can a cognition be of any use to a subject if she has no idea that she has the cognition?[29]

1.4.3 Advaita Vedānta

Whereas the Buddhist reflexivists regard both reflexivity and intentionality to be the distinguishing characteristics of consciousness, the Advaitins regard reflexivity to be the essence of consciousness, intentionality being ultimately illusory. Advaita Vedānta is a metaphysical monism according to which the entire everyday world of distinct objects is ultimately illusory. So, while intentionality is *ultimately* illusory, it is, from the perspective of those in bondage, *no more* illusory than, say, Mt. Everest or even one's own body. In contrast to the world of appearance, pure undifferentiated consciousness (*cit*) is the ultimate reality and this consciousness is self-luminous (*svayaṃjyoti*) or reflexive.[30]

The larger metaphysical framework of Advaita is complex and I will not deal with it here. Hence, my discussion of their reflexivism will be restricted to parts of their accounts of subjectivity, consciousness, and perception. The key concept in Advaita reflexivism is '*sākṣin*', or 'witness-consciousness'. The term '*sākṣin*' means direct or immediate awareness, or the agent of such awareness. It can also denote a witness, both in the legal and in the more general, epistemological sense. *Sākṣin*, as witness-consciousness, is the ultimate witness of all experience, and that which makes

[27] Phillips (2001), p.6 . I have removed Phillips' parenthetical remarks.
[28] See below for further discussion of the regress argument.
[29] Note here that the notion of subconscious mental states is not part of the general context of the debate between Indian reflectionists and reflexivists.

experience possible. Yet, while the concept of *sākṣin* is fundamental to Advaita philosophy, one finds various characterizations of the notion. Gupta takes the following six to be fundamental in the Advaita literature.

1. The *sākṣin* is indubitable, unerring, eternal. It is always directly manifested and its manifestation is not due to any extrinsic reason.
2. The *sākṣin* is the neutral (pure) consciousness as qualified by a modification of the inner sense.
3. The *sākṣin* is that which is never concealed.
4. The *sākṣin* manifests ignorance.
5. That which directly manifests is the *sākṣin.*
6. The *sākṣin* is that which illuminates everything.[31]

Moreover, it is important to note that '*sākṣin*' denotes a *disinterested* witness. The *sākṣin* witnesses all experience, but is not perturbed by it. Hence, it must be strictly distinguished from the narrower empirical ego. The *jīva* (empirical individual) is the 'knower, doer, and enjoyer', bound up in relations with the empirical world. In contrast, witness-consciousness is "completely independent, existing in its own right, and not in relationship to anything else; it is seamless, eternal existence, the ground of our understanding of 'I,' and the ultimate reality that the 'I' names."[32]

So, *sākṣin* is a passive and independent subjectivity, a kind of pure interiority that illuminates everything—including itself—merely by existing. And though it is reflexive in that it manifests itself to itself, in its purest form it does not takes itself as its own *content*, since that would suggest intentionality and perhaps objectification. Yet, despite its ultimate purity and independence, witness-consciousness does figure in the Advaita account of perception and cognition as it occurs in the everyday world.

On the Advaita view, when one is perceptually aware of an object in the world, a representation of the object is presented to witness-consciousness by way of the inner sense (*antaḥkaraṇavṛtti*). In perception, the inner sense literally goes out

---

[30] *Cit* is identical to brahman.
[31] Gupta (1998), pp.3-4 .

through the senses, contacts the object and takes on its form. This formed cognition (what I have called a representation) is then immediately manifest to witness-consciousness because the witness-consciousness is in direct contact with the inner sense. Indeed, on the Advaita view, the inner sense is actually physical (though subtle rather than gross)[33] and acts as a kind of channel for consciousness. The Advaitins give a vivid description of this process:

> Just as the water of a tank, having come out of an aperture, enters a number of fields through channels assuming like those [fields] a quadrangular or any other form, so also the internal organ, which is characterized by light, goes out [of the body] through the door [sense] of sight, and so on, and [after] reaching the location of the object, say a pitcher, it is modified in the form of the objects like a pitcher. This modification [of the internal organ] is called a mental mode [*vṛtti*].[34]

Through this process an object in the world is presented to witness-consciousness. But in addition the cognition of the object is presented to consciousness at the same time. That is, the cognition of the object is self-revealing in the sense that it reveals itself to *sākṣin* in the act of revealing its intentional object. Thus perception involves two forms of reflexivity. The cognition needs no further cognition in order to be revealed to the subject and witness-consciousness itself (which is the subject) is reflexive. Of course, there is a sense in which a perceptual cognition is illuminated not by itself, but by witness-consciousness, since witness-consciousness is the source of all illumination. Nonetheless, a cognition, as a mode of consciousness, is immediately present to *sākṣin*, whereas an object of outer perception is only available to consciousness in virtue of a perceptual cognition being directed toward it. Further, a cognition is a kind of reflection of the object in consciousness and is thus not numerically distinct from witness-consciousness.

---

[32] Gupta (1998), p.17 .
[33] That is, the inner sense is *sattvic.*
[34] Gupta (1998), p. 66

In contrast to perceptual objects, presented to consciousness via the activity of the inner sense and the sense organs, pleasures, pains, and illusory objects are, like cognitions, directly present to witness-consciousness. That is, affective states and illusory objects, such as hallucinations, are not mediated by the activity of the inner sense. This is because these entities are, on the Advaita view, purely subjective. And since they cannot exist if the subject is not aware of them, there does not need to be any mental mode (*vṛtti*) to connect them to consciousness.

Finally, although *sākṣin* is taken to be a pure, disinterested subjectivity, it is also seen as the ground of first-person mental states. As a form of reflexive consciousness, witness-consciousness involves a first-person self-acquaintance that makes possible cognitions with first-person contents. And although not all cognitions need make reference to the 'I', all cognitions are implicitly first-personal in that witness-consciousness accompanies all mental states. Yet, ultimately *sākṣin* is identical with *cit*, the ultimate ground of Being, and as such is not in fact limited to the subjectivity of any empirical individual. Thus, when each of us uses the first-person pronoun we are in fact referring to the same thing! However, since we are not liberated we mistakenly believe that the first-person pronoun refers to the empirical individual (*jīva*).

1.4.4 Kant

No discussion of theories of self-awareness in the western tradition would be complete without a discussion of Kant. However, I should say from the outset that Kant's sophisticated account of the various modes of self-awareness does not fit easily within either the reflectionist or the reflexivist camps. To make matters worse, there are a very wide variety of interpretations of Kant's views on these matters. In

this discussion, however, I will generally stick to the relatively standard interpretation of Kant.

Kant distinguishes between two modes of self-awareness—empirical and transcendental—both of which he takes to be types of apperception. "Consciousness of self according to the determinations of our state in inner perception,"[35] is what he calls empirical apperception. Empirical apperception occurs through a process of self-affection. That is, through self-activity our inner states change and these affect us through being perceived by the inner sense. Thus the inner sense is the means by which the individual intuits her own inner states, whereas the outer sense is the means by which the individual represents to herself objects outside herself. Further, empirical apperception, by means of the inner sense, is consciousness of one's own inner states and is not directly a consciousness of the self. For Kant, empirical apperception, "yields indeed no intuition of the soul itself as an object."[36] However, given that mental states are, for Kant, necessarily states of a subject, empirical apperception is indirectly a form of consciousness of self. In addition, since empirical apperception operates through self-affection, Kant argues that inner sense "represents to consciousness even our own selves only as we appear to ourselves, not as we are in ourselves."[37]

In contrast to empirical apperception, transcendental apperception yields awareness of the subject as persisting through time.[38] Transcendental self-awareness, then, involves neither awareness of one's own states nor awareness of the self as an object. Moreover, of the two modes of self-awareness, transcendental apperception is the more fundamental. Indeed, transcendental apperception is the condition of the possibility of empirical apperception.

[35] Kant (1964), p. A107 .
[36] Kant (1964), p. A22/B37 .
[37] Kant (1964), p. B153 .
[38] Kant (1964), pp. A107, B135, B408-9 .

Kant's account of transcendental self-awareness is based on the relations between consciousness of objects, synthesis, and apperception. The synthesis of a manifold of intuitions constitutively involves the use of the concept of an object in general. Through the use of this concept synthesis produces consciousness of particular objects, and therefore of the objective world. Consciousness of objects, in turn, is related to transcendental apperception through synthesis. It seems that, on Kant's view, transcendental apperception has no content in itself. The subject becomes aware of itself as persisting through time by way of a contrast between the constant variety and change in experience, on the one hand, and its own identity and constancy, on the other. Now, in order for the subject to distinguish itself from the objects of consciousness, the data presented to the sensibility must, through the operation of synthesis, be conceived of as relatively permanent items in a structured, objective world. That is, the subject must be able to distinguish the order of subjective representational states from the order of the objects represented. When the subject is able to make this distinction she is aware of undergoing a structured order of experience within an objective world. Hence it is through the unifying activity of synthesis that a subject is able to be aware of herself and aware of an objective world. On Kant's view, consciousness of self and consciousness of objects arise together.

Now, in characterizing transcendental self-awareness Kant famously remarks that "It must be possible for the 'I think' to accompany all my representations."[39] The *actual* self-ascription of experience involved in making judgments with first-person content is a function of empirical apperception. Moreover, Kant admits that one can have an experience that one does not explicitly ascribe to oneself. Thus, clearly empirical self-awareness is not required for experience. But if transcendental self-awareness is consciousness of oneself as persisting through time and one can have an

[39] Kant (1964), p. B131 .

experience that one does not ascribe to oneself, how are we to understand the role of transcendental self-awareness in this context? There are at least three different ways to understand the relation between transcendental self-awareness and explicit self-ascription. On the first interpretation, experience requires merely the *possibility* of transcendental and empirical self-awareness. That is transcendental self-awareness involves the possibility of empirical self-awareness and experience requires only the possibility of transcendental self-awareness. But this can't be right, for it implies that transcendental self-awareness might not be actualized and, in such a case, empirical self-awareness would be impossible. On the second interpretation, transcendental self-awareness *just is* the possibility of empirical self-awareness.[40] One problem with this interpretation is that Kant not only says that transcendental self-awareness involves the possibility of the 'I think' accompanying all one's representations, he also gives a positive characterization of it—i.e., that it is consciousness of the self as persisting through time. Another problem is that empirical self-awareness requires that the subject grasp her experience as her own, which would seem to require, on Kant's view, a more fundamental sense of oneself as existing through time. That leaves the third possible interpretation: that transcendental self-awareness is an *actual* form of self-representation that makes *possible* (but is not *simply* the possibility of) explicit self-ascription through empirical apperception. On this view, then, all experience involves self-awareness because the operation of synthesis produces both awareness of the self and awareness of object at the same time.

If this interpretation of Kant is correct, then he endorses a form of SAT with regard to transcendental self-awareness, while endorsing a form of DSAT with regard to empirical self-awareness. But is he a reflexivist? My answer is a firm 'yes and no'. Kant holds that all experience constitutively involves self-awareness—transcendental

[40] This seems to be P. F. Strawson's view.

self-awareness is *built into* the structure of experience. Hence, this basic self-awareness is not an act of consciousness distinct from the on-going course of experience. Thus, transcendental self-awareness is not the product of an act of reflection. And so we can say that all experience is reflexive in the sense that it involves not just awareness of objects distinct from itself, but also (and even prior to reflection) awareness of the subject. Moreover, transcendental self-awareness is a form of non-dyadic self-acquaintance similar to the forms of self-acquaintance we have seen in other reflexivist theories. As Kant points out, transcendental self-awareness does not involve awareness of the subject as an object. On the other hand, Kant is clear that transcendental self-awareness is not awareness of one's own mental states. That requires the operation of the inner sense and, according to Kant, one can have an experience of which one is not aware. Hence awareness of one's own states requires an act of reflection, which is a form of inner perception distinct from the states it perceives. So, in the end, it is perhaps most appropriate to say that Kant is reflexivist about transcendental self-awareness and a reflectionist about empirical self-awareness.

1.4.5 Sartre

Some form of reflexivism is a central part of the theories of consciousness of most of the major thinkers in the phenomenological tradition. Brentano argued that every act of consciousness takes a primary object (usually distinct from itself) and a secondary object, which is the very act itself. Husserl argued that all awareness involves an immediate acquaintance with itself and with an ego. Gurwitsch, Zahavi, and Woodruff-Smith all argue for variations on the Husserlian account. But while practically all the major phenomenologists defend reflexivism, only with Sartre do we find reflexivism at the absolute center of an entire philosophy. I will not, however,

discuss the central role of reflexivism for Sartre's account of the mind, his ontology, and his theories of freedom and ethics. My focus here will be only the main structure of his reflexivism.

For Sartre, reflexivity is an essential feature of consciousness. "To be and to be aware of itself," he writes, "are one and the same thing for consciousness."[41] And yet, consciousness is also essentially intentional. Consciousness can only exist insofar as it is directed toward an object, and that object is almost always an object that transcends consciousness. Indeed, like the Naiyāyikas, Sartre holds that consciousness is in itself formless (*nirākāra*), gaining its content from mind-independent objects. Thus, consciousness is "aware of itself in so far as it is consciousness of a transcendent object."[42] However, it is crucial to Sartre's reflexivism that the most basic and pervasive form of self-awareness does not take consciousness as an *object.*

In order to understand Sartre's insistence that consciousness is directed toward a transcendent object, is reflexive, *and* does not take itself as an object, we must begin with two crucial distinctions. First, Sartre distinguishes between *positional* (or *thetic*) and *non-positional* (or *non-thetic*) forms of consciousness. Second, he distinguishes between *reflective* and *prereflective* forms of consciousness.

Sartre holds that every act of consciousness has a positional and a non-positional aspect. The positional aspect of an act of consciousness is its directedness toward an object (much like the objective aspect or *grāhyākāra* in the Buddhist theory). When I am seeing a car pass by outside my window, the car is the posited or intentional object of my visual awareness. Every act of consciousness has some positional object, whether or not that object in fact exists. In addition, "every positional consciousness of an object is at the same time a non-positional

[41] Sartre (1962), p.83 .
[42] Sartre (1962), p.40 .

consciousness of itself."[43] Like the Buddhist notion of a subjective aspect (*grāhakākāra*), this non-positional awareness is awareness of an act of consciousness as directed toward an object. And, of course, as a reflexivist Sartre holds that non-positional awareness is "one with the consciousness of which it is consciousness."[44]

Now if positional consciousness is intentional consciousness and all consciousness is intentional, what is non-positional consciousness? For Sartre, it is a direct, non-intentional, non-dyadic form of acquaintance. The acquaintance referred to here is that involved in, e.g., *feeling* a pain or more generally *undergoing* an experience, and it marks the distinction between a state's merely being *in* a subject and its being something *to* the subject. Further, non-positional awareness does not take its act of consciousness as an object; consciousness is not objectified at the prereflective level. But since non-positional awareness is not an independent awareness in its own right, but is rather an *aspect* of a full blown act of consciousness, the intentionality and the reflexivity of awareness are not in direct conflict. Every full-blown act of consciousness is intentional—that is, it constitutively involves a positional aspect—*and* it is reflexive in that it constitutively involves a non-positional aspect. Indeed, each aspect is impossible without the other. Furthermore, since, on Sartre's view intentionality implies a distinction between subject and object, he holds that basic self-awareness is non-positional. Otherwise, the basic unity of the consciousness act would be disrupted.

When I am aware of an object in the world, that awareness is typically *prereflective* in that its primary focus is something other than itself. Even though I am non-positionally aware of it, my visual awareness of the cup on my kitchen table is focussed on the cup and not my awareness of it. But if I shift my attention from the cup to my awareness of the cup, I now have a *reflective* awareness because the

[43] Sartre (1956), p.iii .
[44] Sartre (1956), p.xxx .

primary focus is now the awareness itself. My reflective consciousness, then, takes my awareness of the cup as a *positional* object insofar as the awareness itself has become the primary focus. Thus, it seems that for Sartre, positional consciousness is a function of attention. And like every positional awareness, a positional awareness of one's own mental acts is accompanied by a non-positional awareness (aspect). That is, reflective consciousness of my awareness of the cup involves a non-positional awareness of my reflective consciousness. Moreover, on Sartre's view, shifting attention from the world to one's own consciousness introduces a certain division in consciousness that is not present at the prereflective level. When I reflect, I attempt to distance myself from myself and turn my consciousness into an object. However, on Sartre's view, consciousness as *being-for-itself* can never be fully objectified. There will always be a remainder that escapes objectification—namely, the non-positional awareness of my act of reflection.

Interestingly, although Sartre defends the self-awareness thesis, he adamantly denies that prereflective self-awareness is a form of self-knowledge. Knowledge, according to Sartre, involves a subject-object division which is absent from our non-positional prereflective self-awareness. Thus, only through reflection can a subject gain self-knowledge, despite her prereflective self-acquaintance. He writes:

> The reduction of consciousness to knowledge in fact involves our introducing into consciousness the subject-object dualism which is typical of knowledge. But if we accept the law of the knower-known dyad, then a third term will be necessary in order for the knower to become known in turn, and we will be faced with this dilemma: Either we stop at any one term of the series — the known, the knower known, the knower known by the knower, etc. In this case, the totality of the phenomenon falls in the unknown; that is we always bump up against a non-self-conscious reflection and a final term. Or else we affirm the necessity of an infinite regress . . . which is absurd.[45]

[45] Sartre (1956), p.xxix .

Reflective self-knowledge, then, must be grounded in a more basic form of prereflective self-acquaintance. For instance, if I am to be explicitly aware that *I* am in pain, it is not enough to be aware that there is a pain; I must also be aware of the pain *as mine*. But how, Sartre will ask, can the act of reflection (if it is itself without even non-positional self-awareness) convey that the pain belongs to the same subjectivity *as itself*? If a third awareness is posited, that only pushes the question back a step. Sartre's solution is to say that reflective self-knowledge presupposes a non-dyadic self-acquaintance and that the act of reflection itself is reflexive.

It is important to note, however, that although prereflective consciousness involves self-acquaintance, it does not involve explicit first-person content. Unlike the Prābhākaras, basic reflexive awareness, for Sartre, does not involve the 'I'. Instead the 'I' enters only with reflective consciousness. Thus the expression of the content of a first-order perception of a pot would not be 'I see a pot', but rather 'There is a pot'. For Sartre, the fact that the person can immediately report her positional awareness of the pot shows that she is non-positionally aware of her awareness; for if she had no awareness at all of her perception, she would not be able to report what she 'perceives'. On Sartre's view, a mental state of which there was not even non-positional awareness would be an *unconscious* mental state, since it would not be anything *to* the subject.

But Sartre does not just hold that 'I' only enters the structure of consciousness with reflection, he further holds that the ego is in fact *constituted* by reflection. For him, there are actually two distinct forms of reflection: pure and impure. Impure reflection involves a double objectification. The reflective consciousness attempts to take the awareness reflected on as an object in much the same manner as an outer-directed positional awareness takes an object in the world. As we have seen, Sartre thinks this move is never really possible because it requires consciousness to take

itself as a transcendent object. Thus it not only attempts to fully divide consciousness from itself; it also attempts to turn consciousness into its ontological opposite, an *in-itself.* In addition, impure reflection involves the objectification of the ego. When impure reflective consciousness attributes a state of consciousness to a subject, it does so in a way that attempts to turn the ego into a *thing* separate from the stream of consciousness.[46] For Sartre, the ego is a mere 'shadow' cast by the for-itself, but impure reflection reifies this shadow and mistakenly substitutes it for the for-itself. Hence, impure reflection is a form of what Sartre calls 'bad faith'.

In contrast, pure reflection, while it involves the introduction of the 'I', does so without the problematic forms of objectification characteristic of impure reflection. But if pure reflection does not objectify the aspect of consciousness reflected on, how does it differ from prereflective non-positional self-awareness? Sartre is by no means clear on the nature of the distinction. One difference of course, is that pure reflection involves an 'I' and an explicit sense of ownership. And the owner here is not the reified ego, but the for-itself. So in pure reflection the subject (or rather, *subjectivity*) is explicit, but not reified, whereas it is only implicit in prereflective self-awareness. However, Sartre also says that pure reflection is a form of knowledge—indeed, pure reflection yields certain knowledge. But if knowledge involves a subject-object duality and pure reflection does not objectify consciousness, then Sartre seems inconsistent. And unfortunately, the best he can do to avoid this inconsistency is to say that pure reflection takes consciousness as a 'quasi-object'. However, despite this unresolved problem, Sartre's account of the reflexivity of consciousness is one of the most sophisticated and powerful in the western tradition.

[46] Again, note the similarity to the Buddhist notion of *ahaṃkāra.*

1.5 Arguments for Reflexivism

Having discussed several versions of reflexivism, I now want to turn to a reconstruction and assessment of several of the key arguments for this position. The arguments to be examined are the memory argument, the reportability argument, the regress argument, the objectification argument, the first-person argument, and the dependence argument. Again, as with the arguments for reflectionism, the arguments do not necessarily form a consistent set.

According to the Buddhist reflexivists, the phenomenon of memory lends credence to their notion of *svasaṃvedana.* When I remember a particular object (or person, event, etc.), I also remember, according to the Buddhists, my *experience* of the object. That is, in recalling a sunset I saw years ago, I am also remembering my *seeing* of the sunset. Moreover, I can in the same act recall both features of the intentional object of the original experience and what it was like for me to experience the object. I can, for instance, remember both the vivid colors of the sunset and, say, a vague melancholia permeating the original experience. On the Buddhist view, this supports reflexivism in that my memory involves a phenomenologically seamless awareness of both object and experience, which in turn suggests that I was aware of both at the time of the original experience. If I was *only* aware of the sunset and not at all aware of my awareness of the sunset at the time, how is that my memory involves an awareness of features of the sunset and the experience?

However, though the memory argument might give a certain phenomenological support for reflexivism, it does not *prove* that reflexivism is the right account of experience. As with most phenomenological arguments of this kind, it is always open to alternative explanations of the phenomena. The Naiyāyika, for instance, could respond that one is aware of both the object and one's awareness of the object because the original experience involved both first-order (*vyavasāya*) and

second-order (*anuvyavasāya*) cognitions. Otherwise, on the Nyāya view, there would be no memory of the sunset to begin with. (On the Nyāya view cognitions which are not apperceived are not reliably recorded in the memory.) Further, it could be argued that the 'seamlessness' with which both object and experience are recalled might be a phenomenological illusion. That is, perception and apperception may *seem* continuous, even though they are ontologically distinct and contingently related.

On the other hand, the reflexivist may respond that one can have a memory involving awareness of both the object and the experience of the object without there being any phenomenological evidence that one was reflecting on one's experience *at the time*. To this the reflectionist cannot respond that one *must* have had a distinct second-order apperception at the time of the original experience, since that would beg the question. Secondly, the reflexivist is unlikely to be impressed by the possibility that the phenomenological seamlessness of the memory is the product of an illusion. For, this would seem to imply that one couldn't tell the difference between reflective and prereflective experience. And this is both implausible in its own right and inconsistent with the reflectionist's own 'phenomenological argument' discussed earlier.

Next, the *reportability argument* is central to the Prābhākara case for reflexivism and is also used by Sartre. The idea here is that one is able report on one's experience immediately, without reflection. For instance, when asked what I am seeing while viewing the sunset, I can report my experience without having to shift my attention away from the sunset. This suggests that I have an immediate, prereflective awareness of my experience, which grounds my ability to report on the experience. In such cases, there does not seem to be the need for an *additional* cognition in order for the individual to report on her experience.

The reportability argument, then, seems to be effective against versions of reflectionism that hold that one is aware of one's experience *only* when one is consciously reflecting on it. However, a reflectionist such as Armstrong could argue that the reportability argument also supports his HOP model of self-awareness. According to Armstrong, consciousness 'in the most interesting sense' involves a kind of automatic apperception that is prior to and more basic than conscious reflection. So the fact that one can immediately report on one's experience without conscious reflection is quite compatible with his view and may even be taken as evidence for automatic apperception. Furthermore, it is not clear that the driver in his 'long-distance truck-driver case' would be able to report on his experience. Hence, we must conclude that while the reportability argument gives some support to reflexivism, it is, like the memory argument, not decisive.

The *regress argument* is perhaps the most commonly employed argument in the reflexivist arsenal. However, the regress argument comes in different forms depending both on the presuppositions of the one employing the argument and the position against which it is used. The point of the argument is to show that, if one is to avoid a vicious regress, one must take at least some states to be reflexive. And if this is so, why not stop the regress before it starts by taking first-order states of consciousness to be reflexive?

Let us start with the regress argument as it applies to HOR theories like those put forward by Armstrong and Rosenthal. Recall that HOR theories attempt to *explain* what makes a state (intransitively) conscious in terms of its being the object of a higher-order representation (either a perception or a thought). Now either the higher-order representation is intransitively conscious or it is not. If it is itself conscious, then it will, according to the HOR theory, require a third-order representation. In this case, the same question can be asked again. And if in turn the third-order representation is

conscious, we are off on a vicious regress. On the other hand, if the HOR is not conscious, then the HOR theorist owes an account of how this representation is able to reveal its object to the subject. How can the HOR reveal its object to the subject if the subject has *no idea* that she has such a representation?[47] After all, the higher-order representation is *unconscious*. In fact, both first- and second-order representations are in themselves unconscious. So how is it that one unconscious state can make another unconscious state conscious?

In response to this type of objection, Rosenthal appeals to his distinction between transitive and intransitive forms of consciousness. He writes:

> This objection disregards the distinction between transitive and intransitive consciousness. HOT's confer intransitive consciousness on the mental states they are about because it is in virtue of those thoughts that we are transitively conscious of those mental states . . . So a HOT can be a source of consciousness for the mental state it is about because the HOT is a transitive state of consciousness; it does not also need to be an intransitive state of consciousness.[48]

The problem with this response, though, is that it is merely a restatement of the view being challenged. Moreover, Rosenthal will have to admit that his notion of transitive consciousness is not what we normally mean by consciousness (which seems to involve intransitive consciousness), since a mental state can be transitively conscious while being (intransitively) unconscious. Indeed, how is a transitively conscious state different from what we would normally call a completely unconscious mental state? The HOR theorists provide no real help here.

The regress argument was also applied to the Nyāya view by the reflexivists in the Indian tradition. However, in the case of the Nyāya theory, it is less clear that the argument is applicable. For the Naiyāyikas, a state's being conscious does not require the subject's being aware of it in any sense. Their theory of *anuvyavasāya* is simply

[47] By 'no idea' here I mean that she is not aware of the state in *any* sense. It is of course possible for me to aware of a state without having the concept of a representation.
[48] Rosenthal (1999), p.743 .

an account of self-awareness and not an account of consciousness *per se.* A first-order awareness reveals its object without revealing itself and a second-order awareness reveals the first-order awareness without revealing itself. Thus there seems to be no vicious regress involved. However, the Nyāya theory must still explain how a state can be conscious even though the subject has absolutely no idea she is in such a state. That is, the Nyāya account of first-order awareness takes it to be a form of transitive consciousness. But, the reflexivist wants to know in what sense we can call such a state conscious if the subject is totally oblivious to it.

The point can be sharpened somewhat by asking whether there would be anything it is like for the subject to be in a state that was merely transitively conscious. It would seem that there is not, since the subject is totally oblivious to the state—that is, the state is *in* the subject, but it is nothing *for* the subject. Yet having a subjective character would seem to be essential to a state's being conscious. Further, suppose that I have a merely transitive perception of a cup in the room. Because the mental state reveals nothing about itself I will not be aware of whether the perception is visual or tactile, since being visual (or tactile) is a property of the representation, not its object. But *conscious* perception isn't like that at all. When I have a conscious visual perception of a cup, I am aware of the cup *as it is presented to me in vision*—indeed, I am aware of what the cup *looks like.* Seeing a cup is just phenomenologically different from touching it. This suggests that I am aware of properties of the cup *and* properties of my representation at the same time, thus allowing me to be aware that I am seeing the cup rather than touching just by having that perception. Hence, the Nyāya view of consciousness is counter-intuitive. Furthermore, if the Naiyāyikas admit that consciousness requires some form of awareness of one's own first-order cognitions—that is, they accept SAT—then either

they will need to embrace an HOR theory of consciousness, in which case they *do* face a regress objection, or they will need to accept some form of reflexivism.

The Advaitins, Kant, and Sartre each argue that certain forms of self-awareness do not involve *objectification* of either the subject or the experience, or both. The idea here is that a perceptual model of self-awareness—according to which a perception (or perception-like) awareness takes another state or the subject as an *object*—is inappropriate because it fails to take into consideration the distinction between subject (or the subjective) and object (the objective). The attempt to take the subject as an object of perception is problematic for these thinkers because it reduces the subject to a mere thing. But the subject, they argue, is precisely that which can never fully become an object. And yet one can be directly aware of oneself, from a first-person (subjective) point of view. Hence, even if there are modes of self-awareness that take the subject as an object, there must also a mode that does not. And this non-objectifying mode of self-awareness is best understood as a non-dyadic form of awareness—i.e., one that does not involve a subject-object structure.

With regard to awareness of one's own states, the argument is that the reflectionist model attempts to turn, not the self, but consciousness into an object for itself. And again, the reflexivist will find this problematic since, the argument goes, consciousness is essentially subjective. Even if consciousness succeeds in objectifying some part or aspect of itself, there will always remain a non-objectified part. Moreover, it seems that a subject can be aware of aspects of her own experience without objectifying them. For instance, I may be aware of certain subjective qualities of my experience such as its overall affective tone without ever turning the affective tone into an object. The affective tone of an experience seems to be the kind of experiential content that one can be non-positionally aware of. However, the reflectionist must argue that to be aware of the affective tone of an experience at all,

we must have a separate awareness that takes the tone as an object. This makes reflectionism less plausible, at least on this score.

What I have called the *'first-person argument'* is actually just a slightly different take on the objectification argument. Suppose that I am aware that *I* have a headache. This is a form of self-awareness in which I am explicitly aware of the headache as mine. Now if coming to be aware of the headache as mine is explained in terms of taking the headache as an object of inner perception, then it will, like all forms of perception, involve criterial identification. That is, I will recognize it as a headache in virtue of its having certain properties, such as being painful and being located in the head. But how will I perceive that it is *my* headache. On a reflectionist model of inner perception, this too will involve criterial identification. The headache might have some property that only I have—perhaps the property of being an object of only my own faculty of inner perception. In any case, the criterial identification of the object will then involve an awareness that *a* (the headache) has some property *F* and that I am (or have) *F*. But this type of awareness requires a prior form of self-awareness in that it already involves first-person content. For instance, I may recognize the headache as mine because it is an object of only *my own* inner perception. But how, the reflexivist will ask, do I know that the inner perception is *mine*? The point here is that objectifying forms of self-awareness are grounded on non-objectifying self-awareness. And of course, this is one of Kant's key points.

Lastly, the *dependence argument* is a challenge to the reflectionist idea that awareness and awareness of one's awareness must be independent. Both Dharmakīrti and Sartre suggest that there are states that depend for their existence on awareness of them. The case is clearest with states such as pain. Can there exist a pain of which the subject is completely unaware? What would such a pain be like? Well, it would be like *nothing at all* to the subject, since she is completely oblivious to it. Yet, on our

normal understanding of states like pleasure and pain, they are at least in part *constituted* by what they are like for the subject. Thus, if we subtract awareness of the pain, it is does not seem there is anything left to call pain. For states such as pain, then, awareness of the state is not *extrinsic* to the state, as reflectionism requires. Moreover, as I will argue at length in Chapter 3, the point generalizes to all states that constitutively involve a subjective feel. However, an individual can surely have a pain without reflecting on or attending to the pain. I might have a light headache all day, even though I rarely stop to reflect on it or attend to it. So if awareness of the pain is intrinsic to the pain and I need not reflect on the pain, this suggests that my awareness is prereflective and non-positional.

To sum up, I think it is fair to say that none of the arguments we have just discussed by itself constitutes a definitive refutation of reflectionism. However, taken together they do show that reflexivism is a plausible and viable position; indeed a position with certain theoretical advantages over reflectionism. In particular, reflexivism coheres with our folk-psychological intuition that a conscious state is one that is something *for* its subject, that our experience is *present* to us in a particularly immediate way. Further, reflexivism is able to avoid either vicious regress or a position that thins out the notion of consciousness as is done by Rosenthal and Nyāya. Reflexivism is also able to account for the basic form of self-acquaintance that anchors first-person content. Moreover, the theory is able to account for forms of self-awareness that do not problematically objectify either consciousness or the subject. Thus, while reflexivism cannot be said to be the clear winner over its rival, it is at this point slightly ahead.

1.6 Conclusion

Although reflexivism is perhaps 'slightly ahead', it is not yet, to my mind, an acceptable theory of self-awareness. First, of course, reflectionism has by no means been decisively eliminated. Second, each of the versions of reflexivism I have discussed so far has been combined with some rather problematic presuppositions. For instance, almost all the reflexivists we have examined hold that the mind is either entirely or to a large extent *transparent* to itself. That is, they hold either that *all* mental states are conscious or that there is nothing in the mind of which the subject is not aware. While this view might have been plausible in the past—indeed, many philosophers in both the Indian and western traditions have taken it to be an almost conceptual truth—in the light of more than a century of research in psychology and related disciplines, it can no longer be maintained. Thus, any viable contemporary model of reflexivism will have to account for the existence of unconscious mental states—mental states to which we have no direct access.

Further, any adequate reflexivism will have to address the issues at the heart of current discussion in the philosophy of mind: self-reference, phenomenal consciousness, and naturalism. No account of self-awareness can avoid addressing the rather large contemporary literature on self-reference. A contemporary reflexivist theory must be able to account for the unique semantics and pragmatics of first-person self-reference, as well as the issue of how a mental state can be self-representing in the way required by reflexivism. In addition, the notion that there is 'something it is like' to have an experience, that the experience has a certain phenomenal quality, brings with it a host of complex issues. And as my discussion of the arguments for reflexivism suggests, these issues are also central to the reflexivist case. Next, some form of naturalism about the mind is at present (and I believe rightly) hegemonic. Thus, a viable reflexivism ought to be at least compatible with

some form of naturalism. Further, it should be empirically *plausible*, at least in the sense that it is not incompatible with established empirical results in the sciences of the mind. Finally, it seems to me that discussions of self-awareness in both the Indian and western traditions have for the most part left out the role of the *body*. I think this is a very serious oversight, for reasons I have discussed in the Preface. Hence, a viable reflexivist theory must account for the (central) role of the body and embodiment in self-awareness.

In the remaining chapters of this dissertation I will construct and defend a contemporary version of reflexivism that draws from both Indian and western sources and attempts to adequately address the complex issues facing the reflexivist account of self-awareness.

CHAPTER 2
THE SEMANTICS OF SELF-AWARENESS

The first-person singular pronoun 'I' is intimately bound up with the phenomenon of self-awareness. Some philosophers have even argued that to be self-aware, at least in the most basic sense, *just is* to have mastered the use of 'I'. I think this is a mistake. To see why it is a mistake, though, will require an examination of the semantics of 'I'-statements. In the process of this examination, I want to make three main points. First, it is a mistake to try to explain (pre-reflective) self-awareness in terms of the mastery of the first-person pronoun because, in fact, such mastery presupposes self-awareness. Second, the semantics of first-person content provide grounds to reject any account of self-awareness based on perceptual or quasi-perceptual self-identification. Third, and finally, a look at the semantics of indexicals more generally—even in statements not involving explicit self-reference—will yield important insights about the connections between the perspectivity of perceptual consciousness, indexical reference, and self-awareness.

We should begin by noting several unique and important features of first-person self-reference.

(1) First-person self-reference is referentially unique—that is, 'I' does not refer in the same way that other referential terms or expressions do.
(2) Proper use of the first-person pronoun *guarantees* self-reference.
(3) This first-person mode of self-reference is not reducible to third-person modes of self-reference.
(4) First-person self-reference requires awareness of self-reference.
(5) Certain forms of first-person self-reference are immune to error through misidentification.

2. 1 'I'

When we say "I am in pain" or "I see a cup," we are (or at least seem to be)[1] referring to ourselves. Indeed, on most accounts of the semantics of the first-person pronoun, self-reference is *guaranteed.* Unlike proper names and definite descriptions— which can referentially misfire either because the purported referent does not exist or because there is more than one referent that fits the name or description—when a speaker uses 'I' correctly the pronoun *cannot* refer to anyone other than the speaker herself. Moreover, the guaranteed reference of 'I' is independent of the truth of the claim being made about oneself. Even when I utter "I am the Queen of England," the first-person pronoun still refers to me, which is why the statement can smoothly manage to be false. Of course, if one accepts that 'I' is a referring term, there arises the difficult issue of what the pronoun in fact refers to, but I will leave aside this more metaphysical question for the time being.

Not only is 'I' referentially unique, it is also irreducible to third-person forms of self-reference. Take, for example, a situation in which I enter my living room and find a trail of blood on the carpet. Upon seeing the blood I exclaim, "Someone is bleeding!" Unbeknownst to me, it is in fact myself that is bleeding. Hence, I have succeeded in referring to myself, but only by a third-person route. Furthermore, the addition of further third-person information will not necessarily lead me to the realization that *I* am the one who is bleeding. Now, in normal circumstances, sufficiently detailed third-person information *can*, of course, lead to first-person knowledge. The point is that piling up third-person descriptions or other such means of identification (e.g., seeing oneself in the mirror) will not *necessarily* lead one to first-person knowledge: it is always possible that I could have some objective piece of

[11] But cf. Anscombe (2000), for the view that 'I' does not refer. See also, Ch. 5 in which I take up the metaphysical issues surrounding this question.

information about myself without *realizing* that it is about me. And this shows that first-person reference is not reducible to third-person reference.

The irreducibility of first-person self-reference is closely tied to the fourth characteristic feature listed above: First-person self-reference requires *awareness* of self-reference. Let us take another example. Susan, a well-known literary personality, is an amnesiac. Not long after waking up to find she has no memory of her own identity, she visits the local library. She picks up the newly published autobiography of a famous literary personality, *Susan: The Whole Story*. Engrossed, she reads the book from cover to cover, never realizing that she is reading about herself. She now knows a great deal about herself and her life, without at all being aware that it is *she herself* that she knows about. If she were to have a conversation with one of her fellow library patrons about the contents of the book she is reading, she would then succeed in referring to herself, all the while unaware that she is telling the story of her *own* life.

Now, let us say that, in the course of her reading, Susan learned that she was born in Cleveland. Obviously, this way of characterizing what Susan learned is ambiguous between first-person and third-person forms of knowledge.[2] The above formulation could express the following:

> (1) Susan learned that Susan was born in Cleveland. (When what she learned would be directly expressed as "Susan was born in Cleveland.")

On the other hand, the formulation could express:

> (2) Susan learned that *she herself* was born in Cleveland. (When what she learned would be directly expressed as "*I* was born in Cleveland.")

Because her case is unusual in that she is an amnesiac, we can conclude that (1) characterizes what she learned from reading her own autobiography. The important

[2] Casteneda (1999).

question, then, is what would Susan need to know in order to have gleaned (2) from her reading rather than (1)? She would need to believe "Susan was born in Cleveland *and I am Susan.*" In order to have this kind of first-person knowledge and to correctly refer to herself by a first-person route, she must be aware that she is referring to herself. Thus we can say that first-person self-reference is an *internal* form of self-reference in that I must be aware that I am referring to myself in order to employ the first-person route to self-reference, whereas third-person self-reference is an *external* form of self-reference in that I may refer to myself through this route without having any awareness that I am referring to myself.[3]

Castaneda marked the distinction between internal and external forms of self-reference by distinguishing the use of the third-person pronoun 'he (him, his)' as a regular reflexive pronoun and as an indirect reflexive, 'he*'. In attributing genuine internal self-reference to another, it is the indirect reflexive that must be employed. Thus if Susan begins to regain her identity, we might say "Susan knows that she* was born in Cleveland," and this specification of the content of her knowledge is directly expressed by Susan as "*I* was born in Cleveland."

So far we have seen that self-reference is guaranteed by the correct use of 'I', that first-person self-reference cannot be reduced to third-person self-reference, and that first-person self-reference requires awareness of self-reference (it is an internal mode of reference). These are features of all forms of first-person self-reference. However, first-person forms of content (either 'I'-thoughts or 'I'-statements) can be divided into two kinds based on their susceptibility to a certain kind of error: error through misidentification. This division can be traced back to Wittgenstein. In *The Blue Book* he writes:

> There are two different cases in the use of the word 'I' (or 'my') which I might call 'the use as object' and 'the use as subject'. Examples of

[3] Zahavi (1999), p. 7

> the first kind of use are these: 'My arm is broken', 'I have grown six inches', 'I have a bump on my forehead', and 'The wind blows my hair about'. Examples of the second kind are: '*I* see so and so', '*I* hear so and so', '*I* try to lift my arm', '*I* think it will rain', '*I* have a toothache'. One can point to the difference between these two categories by saying: The cases of the first category involve the recognition of a particular person, and there is in these cases the possibility of error, or as I should rather put it: The possibility of error has been provided for . . . It is possible that, say in an accident, I should feel pain in my arm, see a broken arm at my side, and think it is mine, when really it is my neighbour's. And I could, looking into a mirror, mistake a bump on his forehead for one on mine. On the other hand, there is no question of recognizing a person when I say I have a toothache. To ask 'are you sure it's *you* who have the pains?' would be nonsensical.[4]

It is important to note that Wittgenstein is not claiming that that the subject-use of 'I' is immune to error *per se*, but rather that this use of the first-person pronoun is immune to the error of *misidentification.* When I am aware that I have a toothache, I do not think that someone or other has a toothache and that someone is *me.* With the subject-use of 'I', the question of identification does not arise. Yet, these forms of first-person content can surely be mistaken in other ways. For example, I might say that I see a pine tree, when in fact the tree I observe is a palm. The point is that, when I say that I see a palm tree, I cannot be mistaken about *who* is seeing a palm tree. This is because my reference to myself in these cases is non-criterial and non-inferential—I do not reason, for instance, that I have the unique properties x, y, and z, whoever it is that is seeing a palm tree also has properties x, y, and z, so I must be seeing the palm.

In contrast, the object-use of 'I' is not immune to error through misidentification. To take Wittgenstein's example, in an accident, I might feel a pain in my arm, see a broken arm and mistakenly believe that the broken arm is my own. Here I believe, correctly, that a is F ('a has a broken arm'), but I mistakenly believe that I am a. Thus we can say that object-uses of 'I' can be analyzed as the conjunction

[4] Wittgenstein (1958), pp. 66-67 .

of two more basic contents: '*a* is F' (a predication component) and 'I am *a*' (an identification component). To return to a previous example, Susan's belief that she was born in Cleveland is not immune to error through misidentification because it can be analyzed into the conjunction of 'Susan was born in Cleveland' and 'I am Susan', both of which involve criterial identification. Whenever criterial identification is involved, there is the possibility of misidentification. Hence we can imagine that our amnesiac is mistaken about who was born in Cleveland because she may hold the false belief that she is Susan. Yet notice that with regard to the 'I' in the second conjunct, reference is still guaranteed. So while guaranteed reference and immunity to error through misidentification may be linked, both object and subject forms of first-person content display guaranteed reference.

At this point one might wonder how we are to understand the content of the identification component 'I am *a*'. Does it involve criterial identification? We can, for example, analyze 'I am Susan' into 'There is a person named Susan' and 'I am that person'. But as Bermúdez points out, if *every* identification component itself involves an identification component, we are off on an infinite regress.[5] We should therefore conclude that "any first-person content subject to error through misidentification will ultimately be anchored in a first-person content immune to error through misidentification."[6] (I will say more on this point below.)

We have seen that what distinguishes the subject-use of 'I' from the object-use of 'I' is the presence in the latter of an identification component—and hence the possibility of misidentification. We should also see, as Gareth Evans has argued, that the subject-use of 'I' is not restricted to the ascription of mental properties. Evans remarks that "Wittgenstein's discussion [in the passage quoted above] does not take sufficient account of the fact that the property of being immune to error through

[5] Bermudez (2000), pp. 7-8 .

misidentification is not one which applies to propositions *simpliciter*, but one which applies only to judgements made upon this or that basis."[7] On Evans' view, what determines whether a form of first-person content is immune to error through misidentification depends on the kind of information which gives rise to the content in a particular instance.

In the example of the car accident, my belief that my arm is broken is open to error through misidentification because of the route by which I gained the relevant information about my body and my immediate environment. In this case, I, perhaps, felt an excruciating pain in my right arm, saw a broken arm close by, noticed the arm was that of a Caucasian male, and concluded that the broken arm was mine and thus the source of my excruciating pain. Notice that much of the information upon which my belief is based derives from visual information about my external environment and the use of criterial identification of the arm in my field of vision. Hence I (mis)identify the arm in much the same way that I might (mis)identify any object in my environment. It is, therefore, because there is an external-object-based identification component to my belief that my arm is broken—and not because the belief involves physical predicates—that it is open to error through misidentification.

Now, if the source of information is changed, the same content (i.e., "My arm is broken") could be immune to error through misidentification. Rather than attempting to match internal information about my body (the excruciating pain) with visual information about my immediate environment (identifying the arm), I might discover that I have a broken arm directly through proprioception. In such a case, I *feel,* not only that an arm is broken, but that *my arm* is broken, rather than *seeing* a broken arm and inferring (correctly or incorrectly) that the arm is mine. When

[6] Bermudez (2000), p. 7 .
[7] Evans (1982), p. 67 .

proprioception is the basis of my belief that my arm is broken, it no longer makes sense to say "someone's arm is broken, but is it mine?" As Evans puts it:

> There just does not appear to be a gap between the subject's having information (or appearing to have information), in the appropriate way, that the property *F* is instantiated, and his having information (or appearing to have information) that *he* is *F*; for him to have, or to appear to have, the information that the property is instantiated just is for it to appear to him that *he* is *F*.[8]

In addition to kinesthetic and proprioceptive bases for knowledge of our own physical properties, Evans also discusses "the way in which we are able to know our position, orientation, and relation to other objects in the world upon the basis of our perception of the world." I know that I am in my living room by recognizing the room and its contents; I know my location in relation to a tree by seeing the tree; I know whether I can fit through a certain space by seeing the space. It makes no sense to say "someone is in my living room, but is it I?" So, again, Evans' point is that subject forms of first-person content are not limited to mental self-ascription. We can have direct, immediate self-knowledge of many of our physical properties as well.

To sum up, we have so far seen that:

(1) first-person self-reference is referentially unique—that is, it does not refer in the way ordinary terms do;
(2) the reference of 'I' is guaranteed;
(3) first-person self-reference is not reducible to third-person modes of reference;
(4) first-person self-reference is an *internal* form of reference—that is, in order to correctly use 'I', one must be aware that one is referring to oneself; and
(5) the subject-use of the first-person pronoun is immune to error through misidentification.

In the next two sections I will develop two important implications of the semantics of first-person content. First, I will argue that mastery of the first-person pronoun

---

[8] Evans (1982), p. 68.

presupposes basic self-awareness. Second, I will argue that, when we take into consideration what is required in order for first-person content to be immune to error through misidentification, we can see that accounts of self-awareness based on the model of what I will call 'object-consciousness' encounter serious difficulties.

2.2 Self-Awareness and the Mastery of the First-Person Pronoun

The abilities both to have thoughts with first-person content and to express those thoughts using the first-person pronoun are obviously central to the phenomena of self-awareness. Indeed, first-person thoughts have first-person contents, and these contents can only be specified in terms of the first-person pronoun or the indirect reflexive pronoun (he* or she*). Could an individual who was incapable of using either of these pronouns have first-person thoughts? Or, put another way, does first-person thought presuppose mastery of the first-person pronoun? Moreover, if mastery of the first-person pronoun is a necessary condition for thinking thoughts with first-person contents, what else might be needed to explain the capacity to think first-person thoughts?

One way to answer this last question is to say that, once we have an account of what it is to master the semantics of the first-person pronoun, we will have explained everything distinctive about the capacity to think 'I'-thoughts—that is, *nothing* further is needed to explain the capacity to think first-person thoughts. We can, following Bermúdez, call this the deflationary account of self-awareness. The basic motivation for the deflationary account is the thought that we should explain what is distinctive about first-person content by an appeal to its immunity to error through misidentification, and that, in turn, immunity to error through misidentification is a function of the semantics of the first-person pronoun. Thus, if we can straightforwardly explain immunity to error through misidentification in terms of the

semantics of the first-person pronoun, then we will have explained everything distinctive about first-person content and the ability to have it.

Bermúdez points out that:

> It is important to stress how well [the deflationary account] meshes with one persuasive strand of thought about self-consciousness. Many philosophers have thought that what is distinctive about beings who are self-conscious is that they are capable of ascribing predicates to themselves. Those predicates are sometimes taken to be psychological (by those of a Cartesian bent) and sometimes both psychological and physical. If it is assumed that the predicates in question have a constant sense, whether they are applied to oneself or to others, then everything distinctive about self-conscious grasp of those predicates that apply to one seems to fall on the act of self-ascription, and in particular, on the first-person pronoun by which that self-ascription is effected.[9]

This line of thought is not far removed from a fully deflationary account of self-awareness. But is such an account viable?

Despite how well it meshes with the above line of thought—and despite its reflecting the general tendency within analytic philosophy to approach forms of mental content, especially thought, by way of language—the deflationary account of self-awareness is not viable. To see why, let us look again at the semantics of the first-person pronoun. The basic rule governing the use of the first-person pronoun is that it necessarily refers to the person using it: as discussed in the previous section, its reference is guaranteed.

However, the problem with characterizing the rule in this way is that it leaves open the possibility that one could refer to oneself without knowing that one has done so. But this won't do because, as I have argued above, first-person self-reference is *internal*: I must know I am referring to myself in order to correctly use the first-person pronoun. The obvious next step, then, is to build this requirement into the semantic rule for the first-person pronoun. We could say that, in employing a token of 'I', a person refers to herself*. By building the indirect reflexive pronoun into the semantic

rule, we close off the possibility that a person could properly use a token of 'I' without knowing that she was referring to herself. But in making this move, a much more serious problem arises for the deflationary account of self-awareness. Recall that the deflationary account is an attempt to explain the capacity to think first-person thoughts in terms of the mastery of the semantics of the first-person pronoun. Now if we build the indirect reflexive into the semantic rule, we run headlong into a vicious explanatory circularity because we cannot account for the indirect reflexive except in terms of the direct reflexive 'I'. This means that the semantic rule will contain first-person content and so cannot be used to *explain* the capacity to think thoughts with first-person content.

A defender of the deflationist account might at this point counter that we need not formulate the rule for the first-person pronoun so as to contain the indirect reflexive. Instead, we could formulate the rule this way:

> If a person employs a token of 'I', then he refers to himself in virtue of being the producer of that token.[10]

Here the use of 'himself' should be taken as the ordinary reflexive pronoun, in contrast to the indirect reflexive himself*. Thus, the explanatory circularity seems to be avoided. But if the indirect reflexive is removed, are we not back to square one, facing again the possibility that a person could employ a token of 'I' without knowing she is referring to herself? This conclusion would be too hasty, though, because the reformulated rule has the speaker referring to himself 'in virtue of being the producer of that token.' This phrase is supposed to do the work previously done by the indirect reflexive. So how does a person refer to herself with 'I' 'in virtue of' being the producer of that token? As we have seen, first-person self-reference is *internal* and so one must know that one is referring to oneself in order to use the pronoun correctly.

---

[9] Bermudez (2000), p. 12 .
[10] Bermudez (2000), p. 15 .

We should, therefore, understand 'in virtue of' to mean 'knowing that'—and knowing that I am referring to myself is a form of first-person thought. Hence, in order to master the semantic rule for the first-person pronoun, it seems an individual must *already* be able to think thoughts with first-person contents. Given the above considerations, the deflationary account of self-awareness seems to be unable to avoid explanatory circularity.

Now, as Bermúdez argues, a second form of circularity lurks nearby. The deflationary account runs into explanatory circularity because it must either employ a semantic rule that itself has first-person content or it must indirectly appeal to a prior capacity for thinking first-person thoughts. If the latter route is taken, then the deflationist is trying to explain the capacity for first-person thought in terms of the capacity to use the first-person pronoun, which, as has been demonstrated, *itself* requires a prior capacity for first-person thought. This is what Bermúdez calls capacity circularity. Now if capacity circularity is understood as just an instance of the interdependence of explanation of two or more capacities, it is not necessarily vicious. Bermudez remarks:

> There is nothing troubling per se about an interdependence of explanation. In the case of self-consciousness, however, problems of circularity are not so easily dismissed, because they . . . [have] implications for the *ontogenesis* of self-consciousness. In brief, if we hold that the various abilities at the root of self-conscious thought form a local holism and consequently can only be explained in terms of each other, then we are ruling out the possibility of explaining how any or all of those abilities can be acquired in the normal course of cognitive development.[11]

If we follow the deflationist approach, then, we will not only rule out from the start the possibility that infants and non-human animals might be able to think thoughts with first-person content (because such thoughts would require language), but we will also be unable to account for how *anyone* could develop these capacities. It seems,

therefore, that some form of self-awareness is required to master the first-person pronoun—that is, use of the first-person pronoun is not the *explanation* of self-awareness, but is rather its *expression.*

2.3 Self-Awareness and Object-Consciousness

In the last section I argued that mastery of the first-person pronoun cannot explain the capacity for self-awareness because this form of linguistic mastery *presupposes* (a perhaps quite minimal form of) self-awareness. In this section I will argue that the semantics of self-awareness raises serious difficulties for any model of self-awareness based on 'object-consciousness'. By 'object-consciousness' I mean any mode of consciousness that reveals what it is conscious of *as an object.*[12] For example, my consciousness of the coffee cup next to my computer presents the cup to me as an object. When I identify or refer to some object, I am identifying or referring to it *as* something or other. Object-consciousness, then, involves the criterial identification of the object of consciousness and is open to error through misidentification—whatever involves identification can involve misidentification.

As we have seen in our investigation of the semantics of first-person self-reference, the subject-use of the first-person is immune to error through misidentification. We have also seen that the immunity to this sort of error is explained by the absence of an identification component in first-person contents (in the subject-use). Suppose, then, that our model of self-awareness is that a form of inner perception (introspection) takes an internal object in much the same way that outer perception takes an object (say, my coffee cup). In this case, both inner and outer perception are forms of object-consciousness, the only difference being what

[11] Bermudez (2000), pp. 18-19 .
[12] We should be careful to differentiate the notion of an object used here from the notion of an intentional object. The intentional object of a mental state is simply what the state is *about* and does

kind of object is perceived. And as a form of object-consciousness, inner perception would be criterial, fallible, and vulnerable to error through misidentification. Let us examine in closer detail why this is the case.

If inner perception is a form of object-consciousness, it will involve object-identification. The inner perception will identify its object as what it is (a toothache, say) on the basis of some criteria or other. This, in itself, is not necessarily a problem.[13] But since inner perception here is a model of *self*-awareness, it must be able to give rise to first-person subject-use contents—inner perception is supposed to explain how I can be aware, not just that there is a toothache, but that *I* have a toothache. But how do I know that that the toothache is *mine*? On the object consciousness model, I would have to be able to identify some characteristic of the toothache that immediately identifies it as being mine. For example, the defender of this view could argue that it is the property of being the private object of only my own introspection (inner perception in this case) that immediately identifies the perceived toothache as my own. However, the immediate question is how I know that the toothache is the object of *my* introspection. At bottom, the problem is that any model of self-awareness based on object-consciousness will reflect the logic of *object-use* of 'I', and the object-use of 'I' is not immune to error through misidentification. Moreover, if we are to avoid an infinite regress, first-person contents involving an identification component must be anchored in first-person contents that do not involve identification—that is, the first-person object-use contents are anchored in first-person subject-use contents. Hence, forms of self-awareness that involve self-identification must be anchored in a form of self-awareness that is identification-free.

---

need to be an object in the sense used here. For example, I might be aware of a *fact*, without holding that facts are objects.

2.4 Indexical Reference and Perspectivity

So far we have discussed contents (mental and linguistic) involving 'I'—what Perry has famously labeled, 'the essential indexical'. And while an understanding of explicit first-person contents is central to understanding self-awareness, there is also a good deal to learn from examining indexical reference more generally. In particular, an examination of indexical reference will allow us to see the close connection between indexical reference and the perspectivity of consciousness.

As Castaneda has pointed out, indexicals express an experiential mode of reference in that indexicals denote items *qua* present in our experience. When I say, 'This pen is out of ink', part of the information conveyed by my utterance is that the pen is directly encountered in my experience. Indeed, it is central to the logic of indexicals that they reveal the subject's orientation to the referent, or if you like, the object's spatial or temporal relation to the subject. In addition, indexical reference is perspectival in that the same objects can be referred to with different indexicals depending on the specific spatio-temporal vantage point occupied by the user. For example, when two people occupy the same room, one person's *here* can be the other's *there*, one person's *this* is the other's *that*. And in order for them to understand one another each needs to grasp not only the layout of the room, but also the various perspectives involved, one's own and that of the other. We can see, then, that indexical reference expresses one's 'perspectival encounter'[14] with objects. Seeing that indexical reference is rooted in a perspectival experiential encounter can help us to understand several of the important features of this type of reference.

---

[13] Unless one holds the, I think unwarranted, Cartesian view that we are never wrong about our own mental states.
[14] Casteneda (1999).

Tomis Kapitan points out five key features of indexical reference relevant to our discussion.[15] First, indexical reference is *ephemeral* due to the changing nature of our perspectives. As I make my way through the world, nothing stays a *this* or a *that*, a *here* or a *there* for very long. Second, no object in the external world is intrinsically indexically positioned; thus being so positioned is an *extrinsic* feature of objects. Third, as with 'the essential indexical', other indexicals are *irreducible*, either to one another, or to non-indexical modes of reference. Fourth, indexical reference is *subjective*:

> One person's "I," "this," or "over there" expresses, in part, what is unique to his or her perspective, making it impossible for another to gain a cognitive fix on the very same item *in precisely the same way*. No one can entertain the exact indexical contents of anyone else, and even one's own indexical contents must differ over time.[16]

Fifth, and finally, the truth-conditions of statements involving indexical reference must reflect the perspectival encounters in which they are rooted. In our discussion of 'I'-statements, it was argued that proper use of 'I' involves an *internal* mode of reference—that is, in order to use 'I' properly I must be aware that I am self-referring. Thus, for instance, we cannot give the truth-conditions of 'I am hungry' (when uttered by me) as " 'I am hungry' is true iff Matt MacKenzie is hungry." This is because the truth-conditions given do not properly reflect the reflexive, internal mode of reference of the statement involving 'I'. Likewise, the truth-conditions of statements involving other indexicals must reflect the perspectival nature of the indexicals. Thus it won't do to say that the truth-conditions of 'This pen is out of ink' can be expressed by " 'This pen is out of ink' is true iff the pen is out of ink."

Now that we have seen that indexical reference is essentially perspectival, the question arises as to the extent of this mode of reference. It is clearly central to our day-to-day lives, and one of the most basic ways for us to 'carve up' our experience.

[15] Kapitan (1999).

MISSING PAGE(S)

71- 91

certain mystery about phenomenal consciousness, and so about phenomenal self-awareness as well.

Despite these caveats, I want to try to dispel *some* of the mystery by giving an account of some processes that might be involved in reflexive awareness. It is unavoidable that this account will be speculative. However, each of the components of the account is defended in the literature on consciousness (which, of course, does not entail that they are correct). My aim here is not to argue that this is how consciousness *must* work, but rather that the reflexivist view is empirically plausible.

3.1.5 Reflexivism and the Integrated Global Representation

I want to begin with the relatively common view that consciousness is an integrated representation, incorporating information from several sense modalities. My experience at any given time (typically) involves information from all of my senses, including interoceptive information.[24] The jury is still out on exactly how this information is selected and integrated—this is the 'binding problem'—but that consciousness involves information from several sources is not particularly controversial. The cognitive neuroscientist Gerd Sommerhoff, for example, takes consciousness to be an *integrated global representation* (IGR).[25] According to Sommerhoff, "the brain forms an extensive internal representation of the current state of the organism which includes representations of the total situation facing the organism both in the *outer* and the *inner* world."[26] On this view, then, consciousness integrates information about the environment and the subject's own states. It has further been conjectured, by Gerald Edelman and others, that the brain engages in

---

[24] Interoceptive information includes proprioceptive information and information from the vestibular system.
[25] Taking consciousness to be an integrated representation is neutral between representationism and phenomenism. Both views agree that consciousness is representational and phenomenal, they just differ in their accounts of how these two aspects are related.
[26] Sommerhoff (2000), p.12 .

extensive blending of internal and external information. The idea here is that certain features of our experience are the result of such blending and that, from the first-person point of view, we cannot always separate the contributions of internal and external sources of information. For instance, it has been suggested that our experience of secondary qualities as 'out there in the objects themselves' is a product of this kind of subpersonal information blending.[27] On the other hand, the IGR also integrates internal and external information in a way that would not necessarily count as 'blending' in Edelman's sense—for example, I am aware of the objects in the room at the same time as I am aware of the position and feel of my body.

Further, it is possible that the IGR incorporates information about itself—information about its contents and/or functional organization—by way of an information feedback mechanism.[28] And it should be pointed out that a representational state can convey information about itself—its representational properties and/or functional relations with other states—in addition to conveying information about its object.[29] As analogies, take, for instance, a sentence and a gauge. The sentence "The cup is blue" conveys information about the cup, but it also conveys information about itself in the process. Since I can read, I am able to glean information about the color of the cup from reading the sentence. But I can also glean that the sentence is in English, that it is four words long, and so on. Likewise, I can glean information about the fuel level in my car by reading my gas gauge, but in the process I also glean information about the gauge itself, such as the placement of the indicator, or the color of the light that comes on when the tank is empty. The important point here is that the representation presents information both about its object and itself at the same time. And since the representation presents information

[27] I am not endorsing this view.
[28] Sommerhoff, Brook, Van Gulick, Edelman, Newton, and others have suggested (though not all have endorsed) a view like this.

about itself, there is no need for a separate second-order representation of the first representation to present that information. What *is* required is sensitivity to both sorts of information.

Now, there is a disanalogy here as well since both the sentence and the gauge are external representations and I employ perception (and other cognitive faculties) to read them. I am not suggesting that we 'read' our representations through internal perception. Rather, I am making the point that in these examples awareness of the properties of the representation is not distinct from awareness of what the representation represents—they are two sides of the same coin. And notice also that the representation does not need to explicitly represent itself in order to present information about itself. This capacity to glean information not just about what a representation represents but also about the representing could be seen as part of the overall function of the IGR: to keep track of (some of) the subject's own states as well as the subject's environment.

Furthermore, taking seriously the idea that consciousness is an *integrated* and *global* representation means that we cannot be overly atomistic about our specification of the contents of consciousness. At any given time, I am aware of many things. I am presently aware of my computer, the sound of a passing car, stiffness in my back, and so forth. But all of these are components of a single, complex state of consciousness. Yet, because, through attention or introspection, I am able to focus on objects in my environment or aspects of my experience, I may talk as if I have many quite distinct awarenesses at any given time—my awareness of the computer, my awareness of my big toe, etc.,—or even as if I am only aware of one object at a time. This is fine, so far as it goes, but one should not forget that the overall character of an experience is determined both by what information it involves and by how that

[29] As an analogy: a sentence conveys information about itself (it's in English, it's 26 words long, etc.) in addition to conveying information about its subject matter.

information is integrated. Moreover, there is a sense in which each representation in the IGR is reflexive in that it reveals to the subject information about both its object and itself. But, on the other hand, one could also say that the IGR itself is reflexive in that it incorporates information about both its objects and its own sub-states.[30]

But why is it important for an organism to have an IGR? The most common and most plausible answer to this is that the contents of the IGR are functionally integrated with a variety of the subject's cognitive and control systems, such as volitional motor control, deliberation, attention, and so forth. Hence, the contents of the IGR are directly available for what David Chalmers calls 'global control'.[31] Indeed, this notion is closely related to the intuitive idea that conscious states are those states that are something *for* their subject. Of course, there could be unconscious states that are available to the subject as well, but these states will not be *directly* or *immediately* available in the way conscious states are: unconscious states are not *salient* to the subject the way conscious states are.

Now, on this account, a state is conscious just in case (or at least, to the degree that) it is a component of the IGR. Notice the difference between this view and HOR accounts of consciousness. On HOR views, an unconscious first-order mental state *becomes* conscious by being represented by a distinct second-order mental state (which isn't conscious either). In contrast, on the IGR view, information conveyed by an unconscious mental state or process is incorporated into the IGR and this process creates a new state, an IGR-state (really, a component or sub-state of the whole IGR). Further, if it is the case that the content of a mental state is in part determined by its functional role within a larger cognitive system, and if we individuate mental states by their contents, then even if we say that an unconscious state *becomes* a conscious

[30] While I will not explore it here, the idea that the IGR is self-monitoring could play an important role in an explanation of the unity of consciousness. Further, Andrew Brook suggests that the IGR (he uses a different term) could itself be the subject of experience.
[31] Block, Flanagan et al. (1999), pp.421-4 .

state, that state will still be significantly transformed in the process in virtue of playing a different functional role. In addition, one of the distinctive things about conscious states, what in part marks them off from unconscious states, is that they have a *phenomenal character*. Thus the idea here is that an unconscious state is either transformed into a P-state (a phenomenal state) by being integrated into the IGR or else the unconscious state gives rise to a P-state in the IGR through information transfer.

With this brief and speculative sketch in place, let me return to the main outlines of PS. First, an experience reveals itself in the act of revealing its object(s). This is accounted for by the general reflexivity of the IGR. The IGR integrates information about the environment with information about the subject, including information about the contents of the IGR itself. Thus the IGR will involve awareness of the positional object and awareness of* the act of revealing that object, including the phenomenal properties of the act.

Second, phenomenal character is not independent of quasi-intentional awareness of that character. On the IGR model, a mental state is a P-state only if it is a state of the IGR and the IGR is such that it keeps track of the states of the subject, including the subject's P-states. This rules out unconscious P-states that need separate second-order mental states to reveal their phenomenal character: no excruciating headaches to which one is oblivious. Of course, the does not entail that the IGR is aware of *all* of the subject's states. It is perfectly compatible with this view to hold that information about some state of bodily damage is, for any number of reasons, not incorporated into the IGR, but that if this information were incorporated in the IGR in the right way the bodily damage would be revealed to the subject as painful.

Third, self-blindness is impossible for P-states. Components of the IGR are directly available for global control (even if to varying degrees), so a mental state

cannot be both an IGR state an inaccessible. Fourth, and finally, phenomenal self-awareness is non-dyadic. This is accounted for because, on the IGR model, there is no separate second-order representation of the IGR. Information about components of the IGR is fed back into the IGR, so the IGR is self-monitoring and its content is self-referential.[32] Thus, this is not a 'Higher Order Awareness' view, but a 'Same Order Awareness' view. Moreover, the self-referential aspect of the IGR is primarily implicit. The thematized or positional objects of awareness are external objects (and the body).

An important implication of the PS model is that phenomenal self-awareness is a more basic form of self-awareness than forms involving explicit first-person content. For example, when specifying the content of my awareness of an external object, I do not need to make explicit reference to myself. (Recall the difference between 'That cup is blue' and 'That cup I see before me is blue'.) Yet, according to the PS model, my experience involves self-awareness. Thus an experience need not be explicitly egological in structure in order for it to constitutively involve self-awareness: I do not need to be explicitly aware of my self *qua* self in order to be, in this minimal sense, self-aware.

And yet, in Chapter 2, I argued that since perceptual experience is inherently indexical, it implicitly involves a form of self-reference. Perceptual experience involves—to use Castaneda's phrase— a 'perspectival encounter' with its object. What this amounts to is that the experiential *mode of presentation* of the object is fundamentally perspectival. Further, as I've argued in this chapter, when we are aware of an object of perception we must also be aware of* the mode of presentation of that object. And since the mode of presentation of the object is essentially perspectival, perception involves awareness of* a perspectival mode of presentation

[32] Keith Lehrer has suggested the certain mental states are self-(re)presenting in that they serve as exemplars of states of that type (including themselves).

of the object. That is, in cases of perspectival experience, phenomenal self-awareness will also be a form of *perspectival self-awareness*, which is an implicit awareness of one's own subjectivity. I should point out that there is more to perspectival self-awareness than is captured in phenomenal self-awareness because perspectival self-awareness in its more robust form involves issues of agency and a grasp of oneself as embedded in the world but distinct from the objects of one's experience. But those issues must be put aside until the next chapter. The task now is to understand the implications of PS for an account of introspection.

3.2 Introspection and Attention

I have argued that phenomenal self-awareness cannot be explained in terms of a dyadic model of the awareness of phenomenal character. But this raises important questions for our understanding of introspection. Even if we don't understand our most basic awareness of phenomenal character in terms of full-blown intentionality, surely when we consciously, volitionally *introspect*, we take the what-it-is-like of an experience as an object of introspection? Well, yes and no. In this section I will argue that we should not take introspection to be a matter of perceptual or quasi-perceptual awareness of the phenomenal character of experience. That is, we should reject an OP model of introspection along with the OP model of phenomenal self-awareness.

I don't deny that we can have consciousness or features of consciousness as intentional objects *in any sense whatsoever.* Of course we can. I can think about the phenomenal character of my experience, wonder about it, write articles about it, glean facts about it, and so forth. What I want to deny is that, when we introspect, we shift our awareness from external objects to internal mental objects or properties, or that introspection involves the operation of a perceptual inner sense such as *manas.* Introspection is not an experience *of* qualia in the same way that seeing is a

perception *of* objects in the environment. Rather, I will argue introspection involves the gleaning of information about the features of our experience. Introspection, then, should be seen not as perceptual awareness-of but awareness-that (or awareness-how). Furthermore, I will argue that introspection depends on our prereflective awareness of* the phenomenal character of experience.[33]

3.2.1 Introspection as Attention to Mental Items

It is at present quite unpopular to hold that there is a distinct sense *organ* responsible for introspection.[34] This is in large part because there seems to be no empirical evidence for an inner sense organ. However, denying the existence of a distinct inner sense organ is not yet to deny the existence of inner sense. One way to cash out this distinction is by arguing that inner sense, at least in part, consists in our ability to attend to certain mental items, such as the phenomenal properties of an experience.[35]

Looking at the scene before me, there are several objects of my awareness: the computer screen, the lamp, the faint hum of the air-conditioner, the light from the florescent bulbs over head, my own body, etc. Moreover, I can shift my attention from one object to another and even attend to them through different sense modalities.[36] But can't I also direct my attention to the phenomenal character of my experience? And doesn't this directing of attention amount to taking the phenomenal properties of the experience—rather than mundane, worldly objects—as objects of awareness? Before taking up these questions directly, however, I need to say something about attention.

---

[33] Several philosophers have developed views of introspection like the one I develop here (Dretske, Shoemaker, etc.). What is important for my purposes here is making explicit the connection between introspection and phenomenal self-awareness.

[34] Though, of course, *manas* was an important sense faculty in classical Indian philosophy of mind.

[35] Attention may not be particularly helpful for an account of introspective awareness of beliefs, thoughts, and the like.

According to William James, attention differentiates awareness into foreground and background.[37] That to which we are attending forms the foreground (including the center of attention), while the background of awareness consists of that to which we are not presently attending. Further, attention can be more or less focussed. When attention is less focussed the distinction between foreground and background will be less sharp. Moreover, since attention structures awareness into foreground and background, it cannot be the case that we are only aware of that to which we are attending. For instance, although I am presently attending to writing and to my computer, I am also aware of a variety of things in the environment, including my own body. Further, I suspect most of us have had the experience of not noticing a constant sound—say, of the air-conditioner or refrigerator—until it stops. In such a case, the sound was part of our background awareness. On this account, then, the shifting of attention can restructure the 'field of awareness' by bringing to the fore what was in the background. We can also sharpen the focus of our attention by attending more closely to an object that was initially part of the foreground of awareness, but was not attended to closely.[38]

So, the question is whether, when introspecting, I am attending to inner items such as mental states, qualia (conceived as either objects or properties), or sensations. To explore this question, I will take up three examples of introspection: introspecting my experience of objects in my environment, an after-image, and a bodily sensation. In each case I will argue that we can understand what it is to introspect these experiences without admitting to an OP model involving attending to inner mental items.

[36] For example, I can attend to the object as it is presented to me visually or tactually.

[37] James (1950), pp.403-4 .

[38] I think James is on the right track here, though, so far as I can see, nothing much will depends on whether James' account of attention is exactly right.

Let's start with a phenomenological experiment, adapted from Shoemaker.[39] Take two objects within your visual field. First, shift your attention from one object to the other. (I find it difficult to do this without also shifting either my eyes or the focus of my vision form one to the other, though perhaps you're more skilled.) Now shift your attention first to the *experience* of the first object and then to the *experience* of the second. Now shift your attention to the *phenomenal character* of each experience.

There must be some phenomenological difference between attending to the object and introspecting. For if there were no phenomenological difference between introspection and *mere looking*, how could a person tell whether or not she is introspecting at any given moment? On the other hand, what precisely is the difference between shifting one's attention between objects in the environment and attending to the experience of those objects? I can detect no visual difference, but this is not surprising since no one claims that inner sense is literally a form of inner *sight.* What, then, is the difference? On Shoemaker's view, ". . . insofar as one does anything different in performing the second attention shift than in performing the first, it will consist in your first *thinking about* your experience of the one [object] and then *thinking about* your experience of the other." And since thinking about an experience is an act of consciousness, there is something it is like to so think.[40] What is gleaned from thinking about one's experience in this case? One gains information about one's own experience, including perhaps information about the phenomenal character of the experience.

After-images are a harder case than the one above. Imagine that a child is playing with two 'glow-sticks', spinning them in circles, one in each hand. It is night,

[39] See Shoemaker (1996), for a slightly more complex version.
[40] It is not necessarily the case that there is something it is like to believe something, because believing that p is a dispositional, rather than an occurrent state. But in so for as thinking is an act of consciousness, there must be something it is like to think a thought. However, as I have said before, unlike perceptual states, we do not specify the content of a thought by what it is like to entertain that thought.

so the after-images left by the glow-sticks are quite clear. It seems that one can easily shift attention between the after-images, so long as they last. And since there is no object out there in the world answering to the bright, florescent green circles, it may seem plausible to suppose that my attention has turned inward to a mental item, a sensation. Indeed, it seems just phenomenologically wrong to say that the attention shift in this case consists in thinking about this after-image and then thinking about the other.

It would be too hasty, however, to say that awareness of an after-image is an awareness of an internal mental item. Nothing in the account of phenomenal consciousness I have given commits me to holding that the object of phenomenal consciousness must be an *existent* object. What is required is an *intentional* object, and intentional objects need not exist. Of course, there needs to be a *cause* of the particular experience, but the cause need not match the intentional object of the experience. A mad scientist might be able to cause me— perhaps by stimulating my brain in some way—to have an experience *as of* a pink elephant, even when there is no such creature in existence. I see no reason not to treat after-images as (non-existent) intentional objects to which one can attend in the same way one attends to other intentional objects. Moreover, since the after-images are part of my visual field, it would be strange to say that my attending to them is a form of introspection, unless one wants to take introspection (or at least one form of it) as a kind of vision. So, since attending to after-images is not significantly different from attending to any other intentional object of perception, the example will not yet count as a case of introspection. Yet if one were to introspect in such a case, it could be accounted for exactly as in the first example.

Finally, let us examine the case of bodily sensations. Suppose I am in a doctor's office, complaining of an upset stomach. It would not be unusual for the

doctor to ask me what kind of pain it is. Pointing to my stomach and remarking that my stomach is the source of my pain is not really what the doctor is after. She wants me to tell her *what it is like* to have this particular stomach ache. So I reflect for a few seconds and respond that it is a sharp, stabbing pain. This is a clear case of introspection in which what I am after is the phenomenal character of my experience. In this case, didn't I just shift my attention from one object to another, from some external object to a pain quale (or qualia)? No, I did not. Rather, I foregrounded my bodily awareness and in so doing I foregrounded the phenomenal character of that awareness along with it.[41] (Recall that, according to the PS model, non-thetic awareness of* the phenomenal character of an experience is not numerically distinct from that experience.) Once I have shifted my attention, I am in a better position to gain information about the experience, including its phenomenal character.[42]

3.2.2 Introspection as Access to Mental Information

What we see in each of these cases is that introspection can be accounted for without accepting the possibility of turning the phenomenal character of an experience into an object of perceptual or quasi-perceptual inner sense. What we gain from introspection is mental *information* (or mental facts),[43] and some of this information will concern the phenomenal character of an experience.

At the beginning of this section, I remarked that introspection depends on phenomenal self-awareness. There are two connected ways in which this is the case.

---

[41] We can think of attention as being used as a kind of phenomenal 'volume control'. By attending to objects in the environment, the phenomenal character of that part of our experience may become more salient, more vivid, etc.

[42] In this example, I've tried to remain neutral between those who think that bodily awareness is intentional (taking certain bodily conditions as objects) and those who do not. Even if bodily awareness is non-intentional, it doesn't affect this account of introspection. However, when we consider the fact that we feel pains as located in different parts of the body, it seems our bodily awareness involves representation. Moreover, it should be noted that whether the information gleaned from bodily awareness involves an identification component is not directly relevant here.

First, without phenomenal self-awareness there could *be* no phenomenal character about which we gain information in introspection. This is because, as we have seen, it makes no sense to separate the what-it-is-like of an experience from awareness of* the what-it-is-like of that experience. Second, when one foregrounds a part of one's field of awareness, one thereby foregrounds the phenomenal character of that part of one's experience.[44] Of course, the phenomenal character does not thereby become an object of attention. Rather, since the phenomenal character (and the awareness of* the phenomenal character) is not numerically distinct from the awareness, it, as it were, 'goes' wherever the awareness as a whole 'goes'. Hence, shifting attention can put us in a better position to gain information about the phenomenal character of an experience.

The reader will have noticed that Shoemaker characterizes introspection as involving thinking about an experience. Yet we shouldn't make too much of this. For example, those who have very strict criteria for what counts as thought—such that human infants and non-human animals lack the conceptual sophistication required for thought—should not on that basis reject this account of introspection. The key point is that introspection involves the gleaning of information about ourselves (our experiences, thoughts, bodily states, etc.). That is, introspection is a form of awareness-that, not a quasi-perceptual awareness-of. Fundamentally, introspection is a form of access-consciousness, whether or not it involves thought in all cases. Furthermore, it is on the basis of the information gained through introspection that we are able to form introspective beliefs.

---

[43] Shoemaker characterizes introspection in terms of mental facts. He does so because introspection is tightly linked to the formation of propositional attitudes about our experience, such as introspective beliefs or judgments.

[44] There are ambiguities in the scope of the term 'awareness' as applied in any given instance. We often talk as if one's present experience is made up of several discrete awarenesses: my awareness of the music, my awareness of the computer, my awareness of my feet, etc. On the other hand, it also makes sense to say that these are all parts of my single present awareness. Likewise with the phenomenal character of an experience. How we specify this will depend on how we specify the experience. Things

Before concluding, I want to point out two further considerations in favor the information-access (IA) model of introspection (in addition to the fact that it avoids the arguments against the OP model.) First, the IA model can make sense of the phenomenological insight that our experience has a certain transparency or diaphanousness.[45] Moore expresses the insight this way:

> [T]he moment we try to fix our attention upon consciousness and to see *what*, distinctively, it is, it seems to vanish: it seems as if we had before us a mere emptiness. When we try to introspect the sensation of blue, all we can see is the blue: the other element is as if it were diaphanous.[46]

When we introspect, we do not focus an 'inner gaze' on mental objects or properties, replacing the objects in the world with inner objects. And since the qualia of an experience cannot be objects of consciousness in this sense[47], if we try to make them objects, they slip away. As Sartre argues, consciousness can never fully become an object for itself. On the other hand, the reflexivist cannot fully endorse the transparency of consciousness. For it is central to the reflexivist account as I am developing it that a conscious state presents information about itself along with information about its intentional object. (If you wear glasses, take them off for a few seconds. Do you glean any information about how your visual perception is presenting your environment?) There is a certain transparency (translucency?) to consciousness, but consciousness is not absolutely transparent.

Second, the IA model meshes well with Evans' insights about the link between information bases and immunity to error through misidentification. The problem with basing first-person contents, such as introspective reports, on an object-consciousness model of introspection such as OP is that it implies an identification

---

are complicated by the fact that how I experience one thing can depend on how that experience fits into my overall experience.

[45] The feature of consciousness is emphasized by the Naiyāyikas, Sartre, and Moore, among others.

[46] Moore (1903), p.450 .

[47] Again nothing I have argued implies that we can't think, wonder, write, etc., about qualia.

component in introspection. And where there is identification, there is the possibility of misidentification. In contrast, on the IA model, introspection *just is* a form of conscious access to the appropriate, identification-free information (e.g., from bodily awareness).

To sum up, I have argued that we should reject not only the OP model of phenomenal self-awareness, but also the OP model of introspection—indeed, many of the same arguments against the former work against the latter. Moreover, the IA model of introspection is not only a natural fit with the PS model, it is a better account of introspection in its own right. However, the account of consciousness I have been developing so far faces serious objections. I will take up these objections in the section after next. I now want to turn to the question of the ego or the (possible) egological structure of consciousness.

3.3 The Ego

On the view of consciousness I have been developing here, all consciousness is self-consciousness in this sense: any awareness constitutively involves non-positional, prereflective awareness of* the phenomenal character of that awareness. I have also argued that, insofar as perceptual consciousness is perspectival, it involves non-positional, prereflective awareness of* the perspectivity of consciousness. These are rather minimal—not to say insignificant—forms of self-awareness. The question I want to address in this section is whether the account of self-awareness I have so far developed implies that consciousness involves, even at its most basic levels, an ego or has an egological structure.

Sartre thought that prereflective consciousness, while constitutively self-conscious, is egoless. Castaneda agrees, reserving the term 'self-awareness' for those experiences which involve explicit self-reference. Thus despite the fact that he holds

that 'unreflective consciousness' involves implicit self-reference, it does not yet count as a form of *self*-awareness.

It is clearly the case that not all experience involves explicit self-reference. Hence, if explicit self-reference is a necessary condition of an experience's having an egological structure, then we should agree with Sartre and Castaneda. But this is too easy. There are two further types of consideration that might lead one to hold that consciousness must have an egological structure. First, one might hold that there must be an ego (or a 'mind's I' or a self) since it is required to account for the unity and identity of consciousness. The idea here is that experience displays a certain unity, a certain coherence that must be accounted for by the unifying function of the ego. Experience has this unity because the experiences all *belong* to a single ego. Moreover, the temporal integration of experience—the fact that what happened to Matt at age 5, happened to *me*, and that my memory of what happened is *my* memory of what happened to *me*—seems to require a temporally identical owner of these experiences. These considerations can lead, and have lead, philosophers to posit an ego outside or behind the stream of experience. Further, these considerations are quite independent of whether all experience involves explicit self-reference.

As will become apparent in Chapter 7, I am not at all convinced by these considerations. It seems to me, though I will not argue for it here, that we can accommodate (or diffuse) these considerations without positing the existence of an ego or self outside of or behind the stream of experience. In particular, I think that we should look, not to an ego, but to the way in which experiences hang together in an embodied stream of consciousness. The subject of experience is the organism as a whole and the unity and coherence of experience is to be accounted for in terms of the way experiences are related to each other and to the organism in which they occur.

Yet, even if one is not convinced by these considerations, the second type of consideration is harder to dismiss. Even if we do not need to posit any such *thing* as an ego, standing outside or behind the stream of experience, it might still be the case that the stream itself (or experience itself) has an egological *structure*. One could argue that the question of the ego is not a question about an entity standing outside experience, but rather a question about the structure of experience—how it is 'given'.[48] In particular, experience is egological, on this view, because it has a first-person mode of givenness. Further, experience involves a first-person mode of givenness even when it is not thematized as one's own (i.e., does not involve explicit self-reference). For instance, when I remember a tree I saw two years ago, the object of my memory is the tree and the content specification of the memory would not necessarily make explicit mention of me.[49] However, according to Husserl and others, there is no question about whose memory it is. The very fact that I can access this memory shows that the memory is mine. And it is in this sense that the memory of the tree has a first-person mode of givenness. Experience has a perspectival structure, and it is in virtue of this structure that we can say that experience is egological.[50]

I have argued in this chapter and the last that experience fundamentally involves phenomenal and perspectival forms of self-awareness and that we have access to information about ourselves and our experience that is immune to error through misidentification. Thus I agree with the details of the above account. On the other hand, by rejecting the notion of an ego outside of experience, to which those experiences belong, one has significantly watered down the contention that experience is fundamentally egological. At this point, one could reasonably reply that

[48] Zahavi reads Husserl this way.

[49] In the case of memory, the subject often is mentioned, but this is not strictly necessary from the first-person point of view.

[50] One might also argue that since experience (at least in any robust sense of the term) involves a grasp of the distinction between the subject of the experience and the contents of the experience, it must

if *that* is all that is meant by 'egological experience' there are few who would dispute the claim. Indeed, this notion of the egological structure of experience is compatible with an experience's being anonymous in Sartre's sense—that is, an experience that involves no thematization of the subject whatsoever. Again, it is not clear to what extent philosophers who take themselves to reject an ego or an egological account of experience will want to disagree. For instance, Owen Flanagan, arguing *against* an egological (or 'egoist', in his terms) conception of experience remarks:

> My consciousness is mine. It is uniquely personal. But this is not because some mind's "I" shadows my experience. It is because of my nature as an organic thinking thing; it is because of the way thoughts hang together in evolved human beings . . . We are egoless.[51]

Flanagan does not deny the importance of the first-person mode of givenness of experience, but he does claim that we are egoless. Rather than claiming that experience must involve an egological structure, we should avoid the philosophical baggage accompanying the term 'egological' and merely insist that experience fundamentally has a subjective, perspectival structure. Moreover, we don't need an ego to account for these structural features of experience, or the fact that certain forms of content are immune to error through misidentification. And since no such ego is needed (or so I contend in Chapter 5), I see no need to use the term 'egological'.

3.4 Objections and Replies

So far, I have argued that, on the one hand, there must be some sense in which we are aware of the phenomenal character of our experience. Yet I have argued that, on the other hand, we cannot explain this awareness on an OP model. The reflexivist, or PS, model is meant to account for our awareness of* the what-it-is-like of experience, while avoiding the problems of OP. I have further argued that an IA

---

therefore have an egological structure. This is an important point that will be taken up in the next chapter. I leave it aside here in order not to overly complicate the present discussion.

[51] Flanagan (1992), p.178 .

account of introspection is both compatible with PS and is independently plausible. I now want to further support my case by answering what I take to be the most important objections to a reflexivist account of self-awareness.

3.4.1 The Long-Distance Truck Driver

In support of his HOP theory of consciousness, D. M. Armstrong describes the case of a long-distance truck driver. He writes:

> After driving for long periods of time, particularly at night, it is possible to 'come to' and realize that for some time past one has been driving without being aware of what it is one has been doing. The coming-to is an alarming experience. It is natural to describe what went on before one came to by saying that during that time one lacked consciousness. Yet, it seems clear that, in the two senses we have so far isolated, consciousness was present . . . That is to say, there was minimal consciousness and perceptual consciousness[52] . . . But something else is lacking: consciousness in the most interesting sense of the word . . . 'introspective' consciousness.

Now if all consciousness involves prereflective self-consciousness, how does one account for this type of case? It seems to be a case of awareness without self-awareness. And this is precisely how Armstrong interprets the case. On Armstrong's view, before the driver came to he lacked a quasi-perceptual introspective awareness of his first-order perceptual awareness of his surroundings. I have already argued that a quasi-perceptual account of introspection is questionable at best. So the question is whether the reflexivist model can account for such cases of 'autopilot'.

There are two responses open to the reflexivist. First, it can be argued that, in fact, the truck driver was not conscious while he was on autopilot. Hence, the example is not really a counter-example to reflexivism. Of course, if one were to take this line, one would need to appeal to non-conscious sensory processing in order to

[52] Recall that, for Armstong, minimal consciousness is consciousness in the sense of being awake or merely having sensation. Perceptual consciousness involves minimal consciousness plus awareness of one's environment.

account for the fact the driver came to while *driving*, and not, say, in a ditch. Perhaps your intuitions differ from mine, but I don't take this response to be very plausible. I have nothing against non-conscious sensory processing; I just see no reason to invoke it in this case. So without some compelling reason, philosophical or empirical, to claim that the driver was unconscious, I think one should grant Armstrong this much. The second response open to the reflexivist is to maintain that the driver is indeed prereflectively self-aware, even on autopilot. This is the response I'll defend here.

Coming-to makes a phenomenological difference in one's experience: there is something it is like to come to after having been on autopilot. Upon coming-to, one becomes more vividly aware of one's surroundings. Further, coming-to may involve explicitly self-referential thoughts such as "I'm driving" or "I've been driving this whole time." Again, Armstrong will explain this phenomenological difference, as well as the possibility of the accompanying thoughts, by claiming that one has become introspectively conscious of one's non-reflexive perceptual awareness.

However, an important question now arises. Was there something it was like to be perceptually aware of the road, etc., while the driver was on autopilot? If there was nothing it was like to be driving on autopilot, then it is not clear in what sense the driver was conscious in the first place. If, on the other hand, there was something it was like, then Armstrong owes us an account of the existence of phenomenal character in the total absence of awareness (reflective or prereflective) of that phenomenal character—one that can get around the objections I have raised so far. Moreover, the reflexivist is not committed to saying that the driver was *reflectively* self-aware even on autopilot. All the reflexivist is committed to saying is that *to the extent that there was something it was like* to drive on autopilot, the driver was prereflectively self-aware. Armstrong's example doesn't rule out this possibility. Furthermore, it is open to the reflexivist to claim that the driver fails to *remember*

what it was like to drive on autopilot, even though his experience did have a phenomenal character—a phenomenal character that cannot exist in the absence of awareness of* that character.

What about the stark phenomenological change upon coming-to? Doesn't that fact show that the driver came *to* self-awareness, rather than being self-aware the whole time? First of all, it is not clear that one needs to explain the phenomenological change primarily in terms of introspection, let alone quasi-perceptual introspection. Rather, the change in the driver's experience can be explained in terms of a shift of *attention*. The reason the driver becomes so much more vividly aware of his surroundings is not that he suddenly (quasi-)perceives his own mental states, but rather that he begins attending to his environment.

There are two different accounts of how this might take place. One might claim that the driver was possibly lost in a daydream or reverie and, in coming to, he shifted his attention back to his driving. The problem with this response is that if the driver were attending to his daydream rather than his driving, he would likely remember that he was daydreaming, if not the content of the daydream. Of course, if we stipulate that the driver remembers nothing from his time on autopilot, we won't know whether or not he was daydreaming. Yet we can easily imagine a case in which the driver was on autopilot, but was not daydreaming. What then?

The second account does not rely on the assumption that the driver must have been daydreaming. Instead, one can say that autopilot is a case of highly diffused or unfocused attention. Attention is often likened to a spotlight in that it can shift from one thing to another. But it is also like a spotlight in that the 'beam' can be more or less focused. Cases of highly focused attention are quite familiar: concentrating on a difficult problem, being absorbed in a film, etc. We can see autopilot as the other end of the continuum. In such cases our attention is not sharply focused on anything.

Hence, coming to involves not the addition of an internal quasi-perception, but rather a focusing of attention on one's surroundings (including, perhaps, one's own place in them).[53] Of course, there is no reason to deny that sharp changes in attention (either shifting or focusing) might be accompanied by (or trigger) introspection. And in this sense, coming-to might lead to reflective self-awareness.

3.4.2 Blindsight

While the phenomenon of 'autopilot' is relatively common in our everyday experience, the phenomenon of blindsight is a much more recent and more surprising discovery. Blindsight was discovered in patients with damage to the striate cortex, the region of the brain that first receives visual input. Lesions to the striate cortex cause blind spots in the patient's visual field. What is surprising about the phenomenon is that patients can discriminate objects within their blind spot with a degree of accuracy significantly above chance. Yet patients deny being able to see anything within the blind spot and often take themselves to be merely guessing in tests of their discriminatory ability. From the patients' perspective, in these experiments, they are not reporting what they perceive, but rather guessing what is in their blind spot—it's just that their 'guesses' are highly accurate.

Blindsight might be seen as a problem for my view in that it could be interpreted as a case of seeing in the absence of awareness of* seeing. Not surprisingly, I don't take blindsight to be a valid counter-example to my view. This is because it is not necessary to interpret blindsight as non-reflexive conscious visual perception rather than unconscious visual discrimination. My view is that every experience involves at least phenomenal self-awareness. Yet I see no compelling reason to count blindsight as an experience. According to the patients' own reports,

[53] James held that we can be aware without any attention whatsoever. Even if we grant this possibility, it does not hurt my case here because we can explain coming-to in terms of *beginning* to attend again

they cannot tell the difference between 'describing' what is in their blind spot and merely guessing. I take this to mean that there is nothing it is like for the patient to 'see' what is in her blind spot, though of course there is something it is like to *have* a blind spot in her visual field. Rather than saying that a patient is having a visual *experience* of which she is unaware, it is at least equally legitimate to say that she has an unconscious visual representation of the object in her blind spot.[54]

An unconscious visual representation, will, of course have a representational character. However, in the case of blindsight, information about this character—about how the representation is representing its object—is unavailable to the subject. Lacking such information, there is no way that this representation *seems* its subject. So, despite the fact that some of the information conveyed by the representation is accessible to the subject, the subject lacks the right kind of information to constitute a phenomenal awareness. Moreover, it seems to me that the information she is able to glean from blindsight is gained through atypical means, insofar as the subject must 'guess' about the object in order to access information about it.

3.4.2 The Distinct Existences Argument

Armstrong argues that introspective judgments, beliefs, and awareness are only contingently connected to their objects because these second-order mental states are the effects of the first-order mental states to which they are intentionally directed. And cause and effect must be distinct existences, only contingently connected.

However, phenomenal and perspectival self-awareness are not second-order effects caused by non-reflexive first-order mental states. Armstrong's argument only shows that if mental state A is caused by mental state B, then A and B must be

after a period in which we were not attending to anything at all.

[54] Sommerhoff holds that we have an unconscious 'Running World Model' which is distinct from the 'Integrated Global Representation' that that he identifies with consciousness. Blindsighted patients may be accessing representations that are part of the RWM rather than the IGR.

distinct existences. This argument might apply to introspective judgments and belief, and acts of reflection generally, but it does not apply to phenomenal and perspectival forms of self-awareness. Yet on the PS model of reflexive awareness, information about the components of the IGR is fed back into the IGR itself, and since this must involve processes in the brain, causation must be involved. But this does not entail a HOP/HOT view. First, the IGR is a system, not a simple, and so I see no reason why a system cannot monitor itself. Second, conscious states (IGR-states) are determined in part by their functional role within the IGR. Third, the IGR can be defined functionally such that its being reflexive in the appropriate sense is an essential feature of it *qua* consciousness. Now if a conscious state is conscious in virtue of its being part of the IGR and the IGR is essentially reflexive, then we will not have the kind of contingent connection between a P-state (recall that only an IGR-state can be a P-state) and awareness of* the P-state (explained by the reflexivity of the IGR) Armstrong wants. That is, phenomenal self-awareness is a form of same-order awareness, not higher-order awareness.

3.4.4 The Intentionality Thesis

According to the intentionality thesis all mental states are intentional. Yet I've argued that phenomenal and perspectival forms of self-awareness are not forms of full-blown intentional awareness. Thus, my form of reflexivism might seem incompatible with the intentionality thesis. However, since my view is also a form of representationalism about consciousness, I can't very well deny the intentionality thesis. The intentionality thesis states that all mental states are intentional. My claim is that states of phenomenal consciousness are reflexive, but I have also claimed that reflexive phenomenal states are intentional. Thus states involving phenomenal self-awareness (reflexive states) are also intentional and so don't contradict the

intentionality thesis. Recall that awareness of* the phenomenal and perspectival character of an experience is not a numerically distinct mental state. Rather, it is an aspect of that intentional experience: an experience reveals aspects of itself *in the act of revealing its intentional object(s).*

Now, one might argue that the intentionality thesis should be interpreted as stating that all mental states are exhausted by their intentionality. But even this stronger reading of the intentionality thesis is not a problem for my view since I hold that phenomenal properties are explainable in terms of representational properties—that is, I account for phenomenal character primarily in terms of representational character. However, if the intentionality thesis is interpreted as entailing object representationalism, then it is incompatible with my version of reflexivism. But in that case, I'm happy to reject it since I think there are independent reasons to reject object representationalism. So, as it stands, my reflexivist account of phenomenal consciousness is not incompatible with the intentionality thesis. All mental states involving phenomenal self-awareness are also full-blown intentional mental states.

3.4.6 Content Specification

Finally, an opponent of reflexivism might argue as follows.[55]

1) Assume reflexivism: a conscious, phenomenal awareness of O is constituted by the same cognizer's awareness of* the awareness of O.
2) Awareness of a blue pot (Abp) = awareness of* awareness of a blue pot [(A)Abp][56] or awareness that one is aware of a blue pot.
3) The awareness that one is aware of a blue pot is not the same as the awareness that one is aware of a pot. With each additional predicate added to the content, the awarenesses themselves change. Hence I am ascribing to myself two different cognitions when I am claiming to have seen a pot and when I am claiming to have seen a blue pot.

[55] I am indebted to Arindam Chakrabarti for making (a slightly different version of) this objection.

[56] '(A)' here symbolizes non-dyadic awareness—awareness-of*—and 'A' symbolizes fully intentional, dyadic awareness, such as perception. Moreover, reflective awareness of awareness of a blue pot would be symbolized: AAbp.

4) Thus Abp is not identical with Ap because otherwise the former (being identical with (A)Abp) would be identical with (A)Ap, which it is not (by 3 above).
5) But we know that a blue pot is a pot because outside of our awareness, being multiply qualified with all sorts of properties does not make it distinct from the same object when any more poorly described.
6) So there is a pot such that whoever is cognizing it as a sheer pot is cognizing the blue pot and whoever is cognizing the blue pot is also cognizing a pot whether they recognize this identity or not. So at least in this one case, Abp = Ap. Between (4) and (6) we are landed in a contradiction. Therefore our assumption, namely reflexivism, must be false.

The gist of this argument, it seems to me, is that externalism about content is incompatible with reflexivism. Reflexivism on this construal is awareness of* the *content* of an intentional awareness. Yet the content of an awareness is at least in part determined by features of an object that may or may not be recognized by the cognizer (as in (6) above). So if I am colorblind and I am aware of the above-mentioned pot, I am aware of a blue pot, even though I am not aware of the pot *as blue.* Thus if externalism about content (even in a moderate form) is true, reflexivism (on this construal) is false.

However, on the reflexivist theory I have been developing, phenomenal self-awareness is not and could not be awareness of *content* in the full externalist sense. Rather it is a non-dyadic awareness of* the *character* of experience. Recall the distinction between the representational character of an experience and the representational content of that experience. The representational character of an experience E is the *mode of presentation* of the object; the representational content of E is determined by what the experience is *of.*

Imagine that Sally is colorblind. She is taken into a windowless room, empty but for a small pot in the middle of the floor. Unbeknownst to Sally, the pot is blue. So, as in (6), her "cognizing it as a sheer pot is cognizing the blue pot." Now imagine that when Sally leaves the room, the pot is painted red. She enters the room again and

perceives the same pot, but this time the awareness is no longer of the blue pot, but is of the red pot. On the externalist view of content, her first and second cognitions had different contents. Yet, the representational *character* of her two cognitions was the same. The representational character is a matter of how the pot is presented to *her*, and since she is color blind, changing the color of the pot won't make any difference to her. In a similar case, imagine that Susan, who is not colorblind, is shown the blue pot. Here the pot is presented to her *as blue*. Yet, upon leaving the room, the first blue pot is switched with an exactly similar blue pot. When Susan comes back into the room, she has another experience with the same character as before (controlling for things like her position, the lighting, etc.). Yet the content of her experience the second time around is different because her experience is not of the same object as the first experience. Again, content varies while character remains constant.

The crucial question is whether phenomenal self-awareness is awareness of* content or of* character. As I have developed the notion of phenomenal self-awareness (and so, reflexivism about awareness) it is a prereflective, non-dyadic awareness of* the *character* of experience. It is an awareness of what it is like to have an experience and of how that experience presents its object (as we have seen, the object is presented *phenomenally*). And as we have seen in the cases of the blue pot and of Inverted Earth, two awarenesses can be phenomenally identical—and therefore identical in representational character—while differing in content. Hence, once the distinction between character and content is in place, my version of reflexivism is insulated from the charge of contradiction.

Changing the symbolism, let, e.g., 'Abp' specify the representational character of the awareness, and let, e.g., 'Abp[bp]' specify the character and content (where the content is specified in brackets). Also assume that the character specification is exhaustive. Further, since phenomenal self-awareness is only of character, let its

scope be set by parentheses. Thus a reflexive awareness of a blue pot *as a blue pot* will be symbolized: ((A)Abp)[bp]. Sally's awareness will then be symbolized: ((A)Ap)[bp]. Susan's awareness will be symbolized: ((A)Abp)[bp]. In Sally's case, the blue pot was painted red, and so her second awareness would be: ((A)Ap)[rp]. Likewise, in Susan's case, the original blue pot was switched with an exactly similar pot. Her awareness would be: ((A)Abp)[bp']. The point of all this is that when the content-character distinction is made, Abp will not equal Ap, even if Abp[bp] and Ap[bp], where '[bp]' denotes the same blue pot. So a reflexive awareness (A)Abp should not be understood as "awareness that one is aware of a blue pot," if this is supposed to include full externalist content. Rather, it is awareness that one has an awareness *as of* a blue pot. And the only way Ap and Abp could specify the same awareness is if the assumption of exhaustive character specification is abandoned.

3.5 Conclusion

The first half of my over-all theory of prereflective self-awareness is now in place. Awareness is reflexive in that it constitutively involves a minimal awareness of* itself. Another way to put this is that consciousness is self-presenting or self-illuminating because it reveals aspects of itself in the act of revealing its object. Moreover, phenomenal self-awareness is not the product of a distinct higher-order act of intentional consciousness directed at our own mental states. Thus, as I've been at pains to show here, the phenomenal character of experience cannot become the object of inner-directed, quasi-perceptual acts of introspection. Hence, I have argued that introspection apprehends information about a subject's mental states. Moreover, the fact that consciousness is reflexive and perspectival is compatible with its being egoless. The task now is to develop the second half of my theory: an account of bodily self-awareness.

CHAPTER 4
BODILY SELF-AWARENESS

In Chapter 1 I said that the body would play an important part in my overall theory of prereflective self-awareness. Yet, so far, not much has been said about the body or the role of embodiment in prereflective self-awareness. In this chapter I want to directly address the crucial role of the body in our most basic forms of self-awareness. As I showed in the last chapter, phenomenal self-awareness is a condition of the possibility of conscious experience, since without it there would not be anything our representing of the world is like *for us.* I will now argue that experience involves self-awareness in another sense: experience of the external world involves[1] awareness of oneself as an embodied being in the world.

My development and defense of this claim will proceed in several stages. First, I will argue that bodily intentionality—reflected or manifested in perception and action—is the most basic form of intentionality. Second, I will then show that bodily intentionality constitutively involves bodily self-awareness in that (a) bodily self-awareness is a kind of 'pivot point' between perception and action, and (b) our perception conveys information not only about the world, but about ourselves. Third, I will discuss the importance of the body and bodily self-awareness for the development of a 'point of view' or subjective perspective. Fourth, and finally, I will argue that prereflective self-awareness is non-conceptual.

4.1 Bodily Self-Awareness

4.1.1 Bodily Intentionality

The first and most basic forms of intentionality are not those of our thoughts or beliefs, but those of perception and action; it is in perceiving and doing that

intentionality is first manifested. Take the action of reaching for and drinking a glass of water. My visual perception of the cup is, of course, intentional in that it is directed toward the cup (its object) and it represents that object in a certain way—e.g., as blue, as containing water, and as within my reach. And while it would be wrong to say that my reaching for the cup represents the cup, the action too is intentional. It is directed toward the object and has conditions of satisfaction (e.g., grabbing the cup in just the right way).

Clearly these forms of intentionality are rooted in our bodily interactions with (and in) the world. Indeed, it is only in virtue of one's on-going, bodily, intentional interaction with the world at the level of perception and action that one could *have* a world about which to think or speak. As Mark Johnson argues:

> In order for us to have meaningful, connected experiences that we can comprehend and reason about, there must be pattern and order to our actions, perceptions, and conceptions. *A schema is a recurrent pattern, shape, and regularity in, or of, these ongoing ordering activities.* These patterns emerge as meaningful structures for us chiefly at the level of our bodily movements through space, our manipulation of objects, and our perceptual interactions.[2]

One need not endorse Johnson's particular approach to see the importance of this point. The world first becomes (experientially) manifest to us at the level of bodily intentionality, as a world through which we make our way.

Yet not just any bodily movement is an action and not just any information up-take counts as perception. There are many creatures that *move* in and *sense* the environment, but do not *act* or *perceive.* One way to approach this distinction is in terms of the appropriate form of explanation of behavior: intentional explanation versus merely mechanistic explanation. If mechanistic explanation is sufficient to

[1] By 'involves' here I mean that awareness of one's own body (as one's own) is incorporated into the overall state of consciousness at any given time. That is, bodily self-awareness is normally an integral part of our integrated global representation (of ourselves and our environment).
[2] Johnson (1987), p.29.

account for the behavior of a creature, then we should not attribute to it perceptions or actions—at least in the particular case at issue.

Consider the case of lab rats that have been conditioned to approach a feeder at the sound of a tone.[3] One way to account for the rats' behavior is in intentional terms. Intentional explanations have three distinguishing characteristics:

> First, they are teleological. That is, they explain an organism's behavior in terms of the purposes and desires that the behavior is intended to satisfy. Second, intentional explanations cannot be eliminated in favor of nonintentional explanations. Third, intentional explanations appeal to desires and purposes in conjunction with nonmotivational representational states.[4]

Concerning the third distinguishing characteristic, two things should be noted at the outset. First, 'representation' should be taken as generally and neutrally as possible. We don't want to be misled by presupposing some particular form of (or account of) representation here. Second, although two types of representation are required for intentional explanation—perceptual representations of how things stand in one's environment and representations of how performing an action can satisfy the relevant desire—a single representation can be of both types at once.

So, if we were to explain the rats' behavior in intentional terms, two types of representation would be involved. A rat would have to have a representation of the goal (the food) and of the instrumental relationship between approaching the feeder and getting food. On the other hand, a mechanistic explanation would make no appeal to a goal or to representation, but would instead rely on law-like connections between stimulus and response. When there are law-like connections between stimulus and response—whether 'hard-wired' or the result of conditioning—there is no reason to posit the kinds of representations relevant to intentional explanation.

---

[3] The example is from Bermudez (2000), p.87 .

[4] Bermudez (2000), p.85 .

Once the rats have been conditioned to approach the feeder upon hearing the tone, an experimenter could then introduce an 'omission schedule'. An omission schedule reverses the relationship between the tone and feeding. Now whenever the rat approaches the feeder at the sound of the tone, it receives no food. If the intentional explanation is the correct one, then the rats should eventually cease approaching at the sound of the tone, since the represented instrumental connection between action and goal no longer holds. However, in such experiments, rats continue to approach at the sound of the tone, despite the fact that they no longer receive food as a result of this behavior. The invariability of this response suggests that mechanistic, stimulus-response explanation is more appropriate here.

The attribution of representation—and the applicability of intentional explanation—is closely tied to the *lack* of law-like connections between sensory input and behavioral output. Certain creatures (including rats in many situations) display a flexibility and plasticity of behavior that cannot be accounted for by purely mechanistic explanation. For instance, with regard to perception certain creatures are able to keep track of distal objects or are able to base behavior on the perceived *absence* of objects (e.g., leaving one's living area because one does not sense a predator in the vicinity). With regard to action, intentional explanation becomes relevant when a creature does something it didn't have to do—that is, the law-like connection between stimulus and response has broken down.

With this brief sketch of the distinction between mechanistic and intentional forms of explanation in place, let me return to bodily intentionality. As I remarked above, I take both perception and action to be forms of bodily intentionality. This is because the body is both the condition of the possibility of action and perception and it profoundly affects their phenomenological structure. Consider once again the example of reaching out and grasping a cup. The movement involved is to be understood teleologically, since it is motivated by the subject's desire to have a drink.

Its role is to contribute to the satisfaction of that desire. It also has more narrowly construed success-conditions involving grasping the cup in the right way (e.g., so as not to allow it to slip out of the hand). Thus, an identical movement in another situation could be a total failure. Moreover, the physiological details (such as, speed, angle, grip, etc.) will make sense only when facts about the cup (its size, shape, weight, distance from the subject, and so forth) are taken into account. It should also be obvious that the subject's behavior is not a mere stimulus response reflex and that it involves representation. But it is important to examine what kind of representation is at work in this case of bodily intentionality.

It would not be unreasonable to assume that the representation of the environment relevant to grasping a cup would involve a representation of the cup in terms of its three-dimensional coordinates in Cartesian space. That is, one might expect that the relevant representation of the cup would be part of a larger, objective or absolute (i.e., non-egological) representation of the environment. However, there is both phenomenological and empirical evidence to suggest that this is not entirely accurate.[5] Rather, the subject represents the cup in terms of its location in egocentric space—that is, in terms of its location relative to her. Further, the grip of the subject's hand is normally set prior to reaching for the cup and is based on her understanding of the relevant physical features of the cup. (It is interesting to note that when one pantomimes reaching out and grasping a cup, the set of the hand is markedly different than when a cup is actually present.[6]) Finally, the action of grasping the cup is typically accomplished with a smooth precision and with rather little in the way of visual feedback, conscious control, or mid-action re-adjustment. This is in marked contrast with the act of locating and pointing out an object. Indeed there is empirical evidence that suggests that grasping and pointing involve significantly different brain

[5] Typical human beings *do* have an objective conception of space. It's just that this conception of space is not what is at work in this case.

processes. Goodale and Milner have hypothesized that much of our bodily intentional behavior involves not primarily 'what' and 'where' representations, but 'what' and 'how' representations.[7] Furthermore, there are cases of dissociative disorders in which patients are able to locate objects in their environment or parts of their own body without trouble when involved in practical activities such as reaching for a pencil, scratching, or dressing. However, when asked to point to or otherwise locate these objects in the objective spatio-temporal order, the patients are unable to do so.

These and other considerations suggest that one's understanding of space at the level of basic bodily intentionality is not objective, but egocentric. Moreover, this understanding of place is not simply egocentric in the sense that one understands the location of objects in terms of one's own location, but also that one understands objects in terms of one's own capacities and projects. And this seems to be the case not just for actions like reaching and grasping, but also for perception. The way we perceive the environment is influenced by past perceptions, expectations, bodily position and orientation, as well as by one's capacities and projects. This is what I meant by saying that the phenomenological structure of perception and action are affected by the body. And these situated, practical and egocentric forms of understanding, perception, and action crucially involve awareness of one's own body.

4.1.2 The Role of Bodily Self-Awareness in Perception and Action

In Chapter 2 we saw that first-person thoughts have an important role in practical reasoning. 'I' is the 'essential indexical' in part because of the way it helps us to make sense of others' behavior. Explanations of the connection between how an individual takes things to be and what she does will often rely on attributing to her thoughts or beliefs with first-person content. The self-awareness reflected in first-

---

[6] Goodale, Jakobson et al. (1994).
[7] Goodale and Milner (1991).

person contents acts as a kind of practical (and explanatory) 'pivot point' between perception and action. The idea is that we will not be able to explain an individual's behavior unless we understand the beliefs underlying it. And we will not be able to understand why those beliefs underlie and motivate the individual's behavior unless we take some of the individual's beliefs to be about herself and her possible courses of action. (Think of the difference between my believing that a hungry tiger is approaching someone or other and my belief that a hungry tiger is approaching *me*.) I contend that bodily self-awareness is the most important 'pivot point' at the level of bodily intentionality. Indeed, bodily self-awareness is a necessary condition of successfully correlated perception and action.[8]

Take, for instance, an individual intentionally walking through a doorway. This action will involve her perception of the doorway as she moves towards it. What does she glean from this perception? What does it contribute to the success of her activity? One crucial thing she must glean from her perception is the location of the doorway in relation to herself. This is important, obviously, because she must come at the opening in the right way in order to make it through successfully. She must also assess the size of the opening in relation to her own size, so that she will know that she can fit through it. The point here is that the correlation between perception and action requires the individual to integrate information about her world with information about herself. Indeed, perception is immediately relevant to action precisely because perception involves information about the world and information about oneself. And at the level of explanation, first-person perceptual contents do the same kind of work as beliefs and thoughts with first-person content.

There are two main sources for the self-specifying information required for the appropriate correlation between perception and action. First, there is interoceptive

[8] It is not, of course, a necessary condition for sensing (or registering) and moving, both of which can be done by creatures who are not conscious.

information, which delivers information about body posture, limb position, movement, and balance.[9] The second source of self-specifying information is from external sense perception. As ecological psychologists have demonstrated, the perceptual field itself carries information about the perceiver. This is particularly true of the visual field. As Gibson points out:

> Vision is *kinesthetic* in that it registers movements of the body just as much as does the muscle-joint-skin system and the inner ear system. Vision picks up both movements of the whole body relative to the ground and movement of a member of the body relative to the whole. Visual kinesthesis goes along with muscular kinesthesis. The doctrine that vision is exteroceptive, that is obtains "external" information only, is simply false. Vision obtains information about *both* the environment and the self.[10]

It is also the case that vision provides information in that there are bodily invariants that circumscribe the field of vision. For example, one's nose is a constant presence in the field of vision. Also, the perceptual array is bounded by the head of the perceiver. Finally, and more controversially, Gibson argues that creatures directly perceive *affordances.* An affordance is a higher level invariant in the visual field partly defined in terms of the capacities and projects of the perceiver: it is something the environment *affords* the perceiver. Gibson gives a particularly instructive example.

> If a terrestrial surface is nearly horizontal (instead of slanted), nearly flat (instead of convex or concave), and sufficiently extended (relative to the size of the animal) and if its substance is rigid (relative to the weight of the animal), then the surface *affords support.* It is a surface of support, and we call it a substratum, ground or floor. It is stand-on-able, permitting an upright position for quadrupeds and bipeds. It is therefore walk-on-able and run-over-able. It is not sink-into-able like a surface of water or a swamp, that is, not for heavy terrestrial animals. Support for water bugs is different.[11]

Thus, both interoceptive and external (exteroceptive) perception provide self-specifying information crucial for successful action in the world. Indeed, without this

[9] Interoceptive information can be divided into proprioceptive information and information from the vestibular organ.
[10] Gibson (1979), p.183 .

integration of self-specifying and other-specifying information, perception would be of little use.

In addition to these Gibsonian considerations, it is important to see the intimate connection between perception, action, and egocentric spatial frameworks. As pointed out above, an egocentric spatial framework is required for the very possibility of action. In order to grasp the cup I must be aware of its location relative to me. Moreover, the egocentric framework structures perceptual experience. I hear a sound coming from *in front* of me and see the hungry tiger *approaching* me to my *left*. Perception has a perspectival structure, and the origin of the perspective is the body. Of course, human beings also have a conception of an objective spatial order. But as Evans and others have argued, having an egocentric spatial framework is a precondition of having an objective (or allocentric) spatial framework. Evans argues that:

> Any thinker who has an idea of an objective spatial world—an idea of a world of objects and phenomena which can be perceived but which are not dependent on being perceived for their existence—must be able to think of his perception of the world as being simultaneously due to his position in the world, and to the condition of the world in that position. The very idea of a perceivable, objective, spatial world brings with it the idea of the subject as being *in* the world, with the course of his perceptions due to his changing position in the world and to the more or less stable way the world is.

Furthermore, perception is not only structured by egocentric *spatial* frameworks, it is also structured by the capacities and projects of the perceiver. This was the point of Gibson's notion of affordances: the perceiver perceives the world partly in terms of what she can *do*.[12]

What emerges from these considerations is the idea that awareness of the world fundamentally involves awareness of ourselves as embodied agents in the

[11] Gibson (1979), p.127 .
[12] It is not only Gibson who insists on this point, of course, but also phenomenologists such as Heidegger and Merleau-Ponty.

world. But before turning to an account of the closely related notion of a 'point of view', I want to give a somewhat more detailed account of the integration of bodily self-awareness with awareness of the external environment.

4.1.3 Bodily Self-Awareness and Awareness of the World

Bodily self-awareness can be explained partly in terms of the *body schema*: the brain's internal representation of the body, including its posture, movements, surface, and so on.[13] The main input to the body schema is from interoceptive information, which includes proprioceptive and vestibular (from the 'balancing' organ in the inner ear) information. Yet the body schema also integrates information from exteroception, especially tactile and visual information. This integration can be demonstrated by closing your eyes, raising your arm, and then using interoceptive information to predict where your arm will be in the visual field when you open your eyes. However, it seems that interoceptive information is dominant for the body schema. For instance, when interoceptive information from a part of a person's body is not integrated into the body schema, that part of the body is felt as foreign object. Despite the fact that the person can see that the body part is connected to *her* body, she will often deny ownership of it. In one experiment, subjects were fitted with inverting prisms that made the world appear upside down. However, the subjects maintained a normal body schema—head up, feet down. It is also important to note that the body schema is usually understood as part of an individual's representation of the *external* world, in virtue of its involving spatial relations.

In the last section I emphasized that bodily intentionality is bound up with an egocentric understanding of space. This can be explained, in part, by the fact that "the body schema and the representation of spatial relations inside the space that can be

[13] Sommerhoff (2000), p.50 .

reached by the arms and hands form an integrated functional unit."[14] The key here seems to be the parietal lobe, which is a cortical region between the visual and somatosensory cortices, involved with both the body schema and the visual control of movement. In one study, cells in the parietal cortex of a monkey fired when a piece of food was within reach, even when the monkey did not reach for it. On the other hand, when the food was just out of reach, no firing occurred. A more striking example again involves 'world-inverting' prisms. When a subject is shown an object outside her reach, the object, as expected, appears upside down. But when the object is brought within the subject's reach, it suddenly appears right side up. These experiments lend support to the idea that bodily self-awareness, bodily intentionality (including capacities for action), and egocentric space are intertwined.

It appears, as several neuroscientists have argued, that the brain integrates interoceptive and exteroceptive information so that this information is available for the successful correlation between perception and action. According to Sommerhoff, "the brain forms an extensive internal representation of the current state of the organism which includes representations of the total situation facing the organism both in the *outer* and the *inner* world."[15] As we saw in the last chapter, he calls this the brain's *integrated global representation* or IGR. On this view, all awareness involves self-awareness in that first-order awareness constitutively involves bodily self-awareness in addition to awareness of the environment. Gerald Edelman conjectures that integration of interoceptive and exteroceptive information occurs through what he calls 'reentrant signaling'. The basic idea is that groups of neurons processing external and internal input exchange information and this integrated or 'blended' information is made available for higher-level processing. He writes, "reentry is a process of temporally ongoing parallel signaling between separate [neural] maps

[14] Sommerhoff (2000), p.50 .
[15] Sommerhoff (2000), p.12 .

along ordered anatomical connections."[16] The result is, again, that first-order awareness involves a blending or integration of awareness of oneself and awareness of the world.

Yet it could be objected that just because the brain blends interoceptive and proprioceptive information, it does not follow that first-order awareness involves bodily self-*awareness*. That is, self-specifying information might remain unconscious. Indeed, it is plausible to think that creatures incapable of consciousness (say, cockroaches) must in some way integrate external and internal information. Moreover, it may be possible to be conscious in the total absence of bodily self-awareness. There are two lines of response I want to give to these objections.

First, there are strong phenomenological considerations that support the idea that our normal experience involves both awareness of the world and bodily self-awareness. I can, if I pay attention, *feel* the position of my body, the pressure on my fingers as I type[17], the location of my left ear, and so on. The ease with which I can move my attention from exteroceptive to interoceptive modalities strongly suggests that interoceptive information is part of my overall awareness, even if it is not usually the focus of my attention. Of course if one identifies consciousness with attention, as some philosophers and scientists have done, then interoceptive information will not usually be conscious. However, I think it is a mistake to identify consciousness with attention.

Furthermore, we can better appreciate the importance of bodily self-awareness by looking at cases in which such awareness is impaired. Oliver Sacks reports the case of Christina, a patient of his who had lost all proprioception. Because she had lost proprioception, Christina was forced to learn to move in the world by relying on vestibular and visual information alone. The process was grueling and she never

[16] Edelman (1989), p.65 .
[17] The feeling of pressure is proprioceptive, not (merely) haptic.

regained the fluidity of movement she once had. In addition, the loss of proprioception produced a radical change in the phenomenal character of her experience. She could no longer feel her body as before; it seemed alien, even dead. The fact that loss of proprioception made such a drastic difference to the phenomenal character of her experience strongly suggests that she lost something that was originally part of her conscious experience: the feel of her body.

But is it the case that bodily self-awareness is a *necessary* condition of consciousness? Anscombe has produced a thought experiment that is meant to show that it is not. She imagines herself in a state of sensory deprivation. "Sight is cut off, I am locally anaesthetized everywhere, perhaps floated in a tank of tepid water; I am unable to speak, or to touch any part of my body with any other."[18] Anscombe holds that in this case her body is not present to her experience, and yet she can think about herself and her situation. Thus she is conscious without having bodily self-awareness. Of course, if Evans is right, she must have a *concept* of herself as an embodied being in the world in order to explicitly think about herself and her situation. But the point here is that, in the present case, she is conscious in the absence of at least phenomenal awareness of her body.

How damaging is Anscombe's thought experiment to my case? Not particularly damaging at all. First, it is not clear what Anscombe's case really shows. If we accept that she has absolutely no bodily awareness whatsoever and that her external senses are deprived of input, what her case shows is that we can imagine that she can still think, that she doesn't lose consciousness entirely. But my claim is that normal consciousness involves bodily self-awareness—arising from the integration of information about oneself and the world—is crucial to the correlation between perception and action required of a successful perceiving agent. Anscombe's case is very much abnormal. And since bodily intentionality is ruled out from the start,

bodily self-awareness is not particularly relevant. As for my claim that the phenomenological structure of perception is shaped by our nature as embodied beings, that is totally unaffected by Anscombe's case. If she did perceive anything, she would do so in a way that is structured by her embodied nature.

My second line of response is primarily to the 'cockroach' objection—that even (presumably unconscious) cockroaches must utilize self-specifying information, so showing that we do too doesn't show that all (normal) awareness involves bodily-self-awareness. And of course, this objection is correct: use of self-specifying information alone does not entail bodily self-awareness. I should point out again, however, that it is not at all clear that cockroaches—or guided missiles, or laptop computers, or a host of other things that use self-specifying information—*perceive* or *act* at all. Bodily self-awareness involves conscious self-representation, but the cockroach probably cannot (it seems to me) be credited with having any representations at all. That is, I suspect that mechanistic explanation is wholly adequate in the case of the cockroach. Thus the cockroach is not a counterexample to my view.

Yet the 'cockroach objection' can lead to another objection: Even if it is granted that self-specifying information is crucial to the correlation between perception and action, the self-specifying information might still be unconscious. In response, it should first be noted that I am not claiming that all self-specifying information is conscious. On the other hand, if self-specifying information is present within the visual field itself, then at least some form of bodily self-awareness is conscious insofar as perception is conscious. Even in the case of the patient who lost all proprioception, she did not lose all bodily self-awareness because she was aware of self-specifying information from vision and from the vestibular system.

[18] Anscombe (2000).

Given these two lines of response to the above objections, I think we are justified in maintaining that normal awareness constitutively involves bodily self-awareness. Yet why call it bodily *self*-awareness, rather than just bodily *awareness*? What makes this form of awareness a genuine case of self-awareness? It cannot be considered a genuine form of self-awareness merely because it makes us aware of our own bodily states. In order for bodily awareness to be a form of self-awareness it must make us aware of our own bodily states (perhaps implicitly) *as our own.* So to know whether bodily awareness is a genuine form of self-awareness, we need to know whether it can yield awareness of bodily states without yielding awareness that the bodily states are one's own states. I might *see* my own body but, failing to see that it is my own body, misidentify it as someone else's. Could I do this in the case of proprioception? It seems that I could not. One of the distinctive characteristics of proprioception is that it only reveals information about one's own bodily states—we're just not wired to proprioceive someone else's bodily states. That is, like introspection, proprioception (and other forms of interoception) is immune to error through misidentification.

Bermudez gives two requirements for genuine cases of self-awareness.[19] First, the awareness "should be about oneself in a way that is nonaccidental." This requirement is met by the fact that interoception is immune to error through misidentification. Second, the content of the awareness must "feed directly and immediately into action." That the contents of bodily awareness are immediately relevant for action can be seen by thinking of the role of bodily awareness in, say, scratching an itch, or in avoiding an object on a collision course with you. Thus bodily self-awareness is genuine self-awareness insofar as it meets these two requirements.

---

[19] Bermudez (2000), p.148 .

Bermudez also distinguishes two forms of self-awareness, both of which involve bodily self-awareness. The narrow form of bodily self-awareness is awareness of one's own bodily properties. The second form of self-awareness is broad self-awareness. "At the core of the notion of broad self-consciousness is the recognition of what developmental psychologists call *self-world dualism.* Any subject properly described as self-conscious must be able to register the distinction between himself and the world."[20] According to Bermudez, bodily self-awareness contributes to broad self-awareness in two ways. First, bodily self-awareness provides awareness of the limits of the body, and this awareness is important for distinguishing (at least) between self and non-self. Second, bodily self-awareness is one of the most fundamental ways in which one becomes aware that the body is responsive to the will. Bodily self-awareness provides the feedback necessary to tell that, for example, one's arm has in fact moved when one tries to move it.

Having taken a closer look at the fundamental features of bodily self-awareness, one can appreciate Merleau-Ponty's emphasis on what he called the 'lived body'—the body in its first-person mode of givenness. I am aware of my own body both as an object in the world and as a subject, from the inside. Indeed, for Merleau-Ponty, the body is both an object and a condition of the possibility of the awareness of objects. And as we have just seen, bodily self-awareness reflects this double nature of the body. It reveals the body as a spatially bounded physical object embedded in a world, and it reveals the body as an intimate presence—something it *feels like* to have (or to be). In addition, bodily self-awareness reveals the body as something that responds to one's will, as that through which one acts. And finally, bodily self-awareness necessarily reveals the body *as one's own.*

[20]Bermudez (2000), p.149 .

4.2 Perspectival Self-Awareness and Points of View

Perspectival self-awareness is a subject's (prereflective) awareness of her own perspective, her own point of view on and in the world. This form of self-awareness straddles both phenomenal and bodily forms of self-awareness. It is involved in phenomenal self-awareness in that phenomenal self-awareness includes awareness of* the phenomenal and representational character of the subject's experience; and insofar as experience is perspectival, phenomenal self-awareness will involve perspectival self-awareness. Perspectival self-awareness is bound up with bodily self-awareness in that what it is to *have* a perspective, to have a point of view, involves both the body and bodily self-awareness.

Crucial to an individual's having a perspective is that individual's ability to distinguish between her experience and what it is an experience of. (And it should be noted that this ability is crucial to full-blown intentionality as well.) Strawson explains this requirement as follows:

> A series of experiences satisfying [this] Kantian provision has a certain double aspect. On the one hand it cumulatively builds up a picture of the world in which objects and happenings (with their particular characteristics) are presented as possessing an objective order, an order which is logically independent of any particular experiential route through the world. On the other hand it possesses its own order as a series of experiences of objects. If we thought of such a series of experiences as continuously articulated in a series of detailed judgements, then, taking their order and content together, those judgements would be such as to yield, on the one hand, a (potential) description of an objective world and on the other the chart of the course of a single subjective experience of that world.[21]

So having a perspective constitutively involves being able to distinguish the 'subjective route' of one's experience from the objective world through which it is a route.[22] Perspectival consciousness, then, is necessarily 'nonsolipsistic consciousness'. Moreover, Strawson's notion of a subjective route also requires that

---

[21] Strawson (1966), p.105 .
[22] Strawson (1966), p.104 .

the individual's point of view be temporally extended—that is, that the distinction between experience and the world is maintained through time.

Now Strawson thinks that the ability to make the relevant distinctions requires concepts. But nothing said so far requires one to follow him on this, so I will withhold judgment on that question for now. Even if it turns out that concepts *are* required in order to have a perspective—though I will argue in the next section that they are not—it will be helpful to formulate the notion of a perspective in a way that does not build in a requirement of concept-possession. Thus, for the purpose of this discussion I will use Bermudez's characterization of a perspective. He writes:

> [H]aving a temporally extended point of view on the world involves taking a particular route through space-time in such a way that one's perception of the world is informed by an awareness that one is taking such a route, where such an awareness requires being able to distinguish over time between subjective experience and what it is experience of.[23]

The notion of bodily self-awareness articulated in the last section goes a long way toward explaining the development of a perspective and perspectival self-awareness. The self-specifying information available in the visual field is to a large extent information about a subject's spatio-temporal route through the world. Thus there is an important sense in which bodily self-awareness is a continuous awareness of oneself as taking a route through the world. In addition, proprioceptive and tactile awareness—which involves both proprioceptive and exteroceptive elements—allow a subject to make a primitive, but crucial distinction between self and non-self.[24] Finally, the continuous and relatively stable—stable in comparison to the constantly changing contents of the subject's awareness of the world—background awareness of one's own body (potentially) contributes to a sense of self through time.

---

[23] Bermudez (2000), p.168 .
[24] However, the self/non-self distinction is significantly less sophisticated than the self/*world* distinction.

The problem is that we can imagine a creature that has bodily self-awareness (in some primitive sense), but whose experience is limited to the continuous present. Having a perspective requires that the creature in some sense grasp that the object of experience is independent of that particular experience—that the object either has existed in the past or will exist in the future. Yet a creature locked into the present cannot have such a grasp—something more is needed.

Clearly, a grasp of basic temporal concepts would do the trick, but it is worthwhile to ask whether the mastery of concepts is necessary to free a creature from the continuous present. A creature without concepts may still escape the continuous present if it has certain basic recognitional capacities.[25] For *re*cognition is bound up with an awareness that what is recognized has an existence that goes beyond the present moment. Yet, while such a capacity would free the creature from the continuous present, the capacity to recognize something is not sufficient for a grasp of the distinction between experience and the object of experience. This is because the creature might only be able to recognize aspects of its own experience, such as pain or hunger.

Now if a creature could recognize physical objects, that ability would be sufficient for the distinction between experience and its object. But an understanding of a physical object *qua* physical object seems to involve conceptual abilities. Alternatively, one might suggest that what the creature must be able to recognize are not physical objects, but places. The idea here is that the fundamentally spatial structure of the experienced environment "provides the basic material for the exercise of basic recognitional capacities."[26] The problem, however, is that the identification of places and the identification of objects seems fundamentally intertwined. Obviously, our notion of a physical object involves spatial concepts, for instance in that the

[25] On some accounts of concepts, any recognitional capacity would count as a concept. I take the more conservative view that only certain rather sophisticated recognitional capacities are conceptual.

criteria of identity (and reidentification) for such an object will involve location and spatio-temporal continuity. It could also be argued that the identification of places will depend on the identification of objects.

Yet it may be that, while our concepts of object and place are bound together, a creature lacking our robust concept (or any concept) of an object might still be able to reidentify places. Bermudez offers two suggestions.[27] First, much recent work in developmental psychology suggests that infants parse their visual perceptions into bounded segments and that the infants have certain expectations about the behavior of those bounded segments. However, perceptual awareness of bounded segments in one's visual array is not yet to perceive objects in the full sense, much less to have mastered the concept of an object. So the suggestion is that the ability of a creature to recognize places might rely on these bounded segments, rather than the perception of objects.

The second suggestion involves the recognition of certain features of the environment, such as warmth, food, or danger. This kind of recognition is different from the recognition of objects (or even quasi-objects, such as bounded segments of the visual array) in that features are not predicated of particular objects—e.g., 'it is raining', 'it is warm', 'there is water', etc. Now a creature that can perceive such features and can navigate on the basis of them might also be able to reidentify places on the basis of such features. That is, the creature might be able to recognize 'the warm place', 'the sleeping place', 'the water place', and so on. But here we must be careful because there are two ways of construing recognition or reidentification. In the weak sense, a creature can reidentify a place if it can find its way back to it. It will, of course, need to be sensitive to certain environmental features, such a particular scent, in order to find its way back. However, the weak or implicit form of

---

[26] Bermudez (2000), p.171 .
[27] Bermudez (2000), p.173 .

reidentification does not yet involve conscious awareness on the part of the creature that it has been there before. Strong or explicit reidentification will involve a conscious awareness on the part of the creature that it has been there before. And what this requires is *conscious* memory.

When a creature can remember having been in a place before, or can remember a quasi-object or feature, it has the beginnings of a robust perspective. To be able consciously or explicitly to reidentify a place is to have some understanding of oneself as moving through space over time. Moreover, as Bermudez points out, "the richness of the self-awareness that accompanies the capacity to distinguish the self from the environment is directly proportionate to the richness of the awareness of the environment from which the self is being distinguished."[28] What explicit place reidentification allows for is an understanding of the environment as composed of items—quasi-objects, features, or places—whose existence goes beyond the present moment. And this understanding is a necessary condition of possessing a temporally extended perspective.

As pointed out above, in addition to nonsolipsistic consciousness, a creature with a perspective must have some awareness that it is taking a particular route through space-time. And this awareness is linked to the creature's navigational abilities. Recognizing places is inextricably bound up with navigation through the world. Moreover, this awareness will be (at least implicitly) self-referential in that the awareness is awareness of the creature's own route through space-time. Recall, however, that there are two basic criteria that this form of awareness must meet in order to be considered a genuine case of first-person content. First, the awareness cannot be merely accidentally self-referential. Second, the contents of the awareness must feed directly into action. A creature's awareness of its route through space-time will meet the first criterion because it is necessarily based on conscious registering of

self-specifying information. This kind of awareness cannot be accidentally about oneself. (Contrast this type of awareness with a creature seeing the reflection of a moving creature in a mirror. The perceiver might gain information about the reflected creature's route and it might turn out the perceiver *is* the reflected creature. This would be accidental awareness of its own route.) As for the second criterion, a creature's awareness of its own space-time route, being inextricably bound up with its navigational abilities, is clearly relevant to action in the required sense.

It has been argued by philosophers and psychologists working on navigation and spatial reasoning that a creature's awareness of its own route through space-time requires a minimal understanding of the nature of space. This understanding of space plausibly consists in the following abilities:

- the capacity to think about different routes through (and to) the same place;
- the ability to track changes in the spatial relations between things caused by it's own movements;
- the capacity to think about places independently of particular quasi-objects or features located in the place.

Furthermore, it is argued that we can explain these abilities in terms of a creature possessing an integrated representation of the environment over time.

While it would take me too far afield to discuss the intricacies of this kind of representation, the important point is that having a temporally extended perspective involves both awareness of oneself and an awareness of the environment as composed of items to be encountered and places in which one can be located and through which one can navigate. So nonsolipsistic consciousness and spatial awareness are bound together in that a creature can understand its distinctness from its environment only if it has a (relatively rich) understanding of an independent environment. Further, this understanding of the environment constitutively involves an understanding of the spatial features of that environment. Moreover, we can see that bodily self-awareness

[28] Bermudez (2000), p.178 .

is closely intertwined with having a perspective. Bodily self-awareness is a crucial source of self-specifying information: through it one is aware of the distinction between self and non-self, and it is necessary (though not sufficient) for keeping track of both one's movement and the effects of one's action on the environment. In addition, one's most basic understanding of space relies on an egocentric spatial framework, which itself relies on both the fact of embodiment and bodily self-awareness. Finally, given the above considerations, we can see that having a perspective is inextricably bound up with prereflective self-awareness, both phenomenal and bodily.

4.3 Non-Conceptual Self-Awareness

4.3.1 Non-Conceptual Content

In previous sections I have been careful not to presuppose that either bodily self-awareness or the possession of a perspective require concepts. I now want to take up the general question whether prereflective self-awareness—phenomenal, perspectival, or bodily—must involve concepts.

The idea that certain mental states have non-conceptual content has received a great deal of attention recently. The usual candidates for being a state with non-conceptual content are perceptions and subpersonal mental states. I will put aside questions about subpersonal states here. But let me begin by giving a general definition of non-conceptual content.

> For any state with content, *S*, *S* has non-conceptual content, *P*, iff a subject *X*'s being in *S* does not entail that *X* possesses the concepts that canonically characterize *P*.

The first, and most obvious, thing to notice about this definition is that it covers only states with *content*. That is, only representational states can count as

having non-conceptual content.[29] Secondly, one can be in a state with non-conceptual content (hereafter, NC-state) even if one *does* have the concepts required to give a canonical characterization of that state. Moreover, it is an open question whether subjects that have no concepts at all can be in NC-states. Thirdly, the canonical characterization of a state is one that specifies the content of that state. For instance, take a perception of a red apple. If this perception is an NC-state, then being in this state (seeing the red apple) does not presuppose that the subject possesses the concepts *red* and *apple*, even though one would employ these concepts to specify the content of that state.

What, then, are the motivations for positing NC-states? First, the richness and fineness of grain of perceptions seems to outstrip our conceptual repertoire. For instance, it seems plausible that we can discriminate more colors or shades than we have concepts for. One way to capture this is to say that perceptual experience has an 'analog' character. As Peacocke makes the point:

> To say that the type of content in question has an analogue character is to make the following point. There are many dimensions—hue, shape, size, direction—such that any value on that dimension may enter the fine-grained content of an experience. In particular, an experience is not restricted in its range of possible contents to those points or ranges picked out by concepts—*red, square, straight ahead*—possessed by the perceiver.

The surface of my kitchen table has many different colors and shades. I am able to perceive these shades and colors, but does this mean I have a *concept* for each shade I can distinguish? I am fairly certain that I could not reidentify every shade I am presently seeing. But if I cannot reidentify the shade, it seems too much to say that I nonetheless have a concept of it. And even if I did have such a concept, is it in virtue of possessing the concept that I am able to perceive the shade in question? When I enter a room full of abstract sculptures, I see things in the room as having particular

---

[29] By 'representational' I just mean that the state is intentional; it has content and construes its object as being thus-and-so.

shapes; but it seems implausible to assume that this requires my having in advance concepts of the particular shapes I am seeing.[30]

Second, one may take perceptions (or other states) to be non-conceptual because they are not linked to other states in the way typical conceptual states such as beliefs are. The idea here is that conceptual states such as beliefs are linked with other conceptual states (C-states) through a variety of 'normative liaisons', such as inferential connections. Moreover, the specification of the constituents of C-states necessarily involves these liaisons, whereas the specification of the constituents of NC-states will not necessarily make reference to such liaisons. For example, if a subject believes that *a* is an apple, and believes that some other object *b* is an apple, then she must believe that there are at least two apples. If she believes that the apple is nutritious, then she must also believe that it is edible. Propositional attitudes are holistic in that, if a subject has one propositional attitude, she must have many. And the conceptual content of these attitudes will be (at least in part) specified in terms of these kinds of rational connections. Likewise, part of what it is to possess a certain concept is to have dispositions to make and to accept as legitimate certain inferential moves between propositional attitudes.

In contrast, the constituents of NC-states will not (or not necessarily) be specified in this way. For instance, it is hard to see what sense could be made of inferences between perceptions (rather than perceptual beliefs). Further, perceptions are not holistic in the way beliefs are. The content of a perception does not by itself entail that there are other perceptions that the subject *ought* to have: perceptions don't *entail* anything. The subject might, on the basis of some perception, come to expect some future perception, but that is another matter. Moreover, philosophers such as Tim Crane[31] have argued that perceptual contents can be contradictory in a way that is

[30] Peacocke (1986), p.15 .
[31] Crane (1992).

ruled out in the case of beliefs.[32] Finally, perceptions are not revisable in the way beliefs are. I may come to believe that what I see is an optical illusion, but what I am presented visually does not thereby change.

Third, and finally, one may posit NC-states in order to explain the possession or development of concepts. It is plausible to explain the acquisition and possession of, say, color concepts partly in terms of the capacity to respond to color perceptions in the right way. But one courts circularity if one attempts to explain the content of those perceptions and responses in terms of color concepts. Further, if one does not accept that there are innate concepts, then there is a strong pull toward ascribing non-conceptual representational states to infants as part of an account of the acquisition of concepts. For instance, if one accepts a representationalist account of consciousness, it is plausible to hold that an infant could visually represent an object as red, without yet having the color concept *red*. Furthermore, it is plausible to appeal to the infant's ability to visually represent color (to see in color) in an explanation of the infant's development of color concepts. In addition, I should point out that even if one accepts that human infants are born with some concepts, it is still plausible to hold that their representational capacities outrun their conceptual capacities. (Again, an infant might be able to see—i.e., visually represent—magenta without having the concept of magenta.)

Having sketched some of the considerations in favor of non-conceptualism, I now want to examine whether the forms of prereflective self-awareness I have discussed might plausibly be considered non-conceptual. Along the way I will consider—and reject—some important objections to non-conceptualism about prereflective self-awareness.

---

[32] It is controversial whether holding contradictory beliefs is ruled out, and if so, in what way. However, one could perhaps weaken Crane's point here by arguing that the normative connections between beliefs make simultaneously and explicitly holding contradictory beliefs more cognitively dissonant than having contradictory perceptions.

4.3.2 Non-Conceptual Prereflective Self-Awareness

Phenomenal self-awareness is a form of immediate, non-inferential awareness of one's own experience. It is a prereflective awareness that one is aware of the world as well as an awareness of how one is aware of the world. Does this kind of awareness require concepts? If so, what kind of concepts? Since prereflective self-awareness involves implicit first-person contents—we are aware of our own experience at least implicitly as our own—one might argue that prereflective self-awareness requires a concept of oneself. If one accepts this view, there is a strong pull toward positing an innate self-concept. If phenomenal experience requires prereflective self-awareness and prereflective self-awareness requires a self-concept, then phenomenal experience requires a self-concept. If this self-concept is not innate, one would be forced either to argue that it develops *very* early or that phenomenal consciousness develops later than is usually thought. Further, as I have argued in Chapter 2, mastery of the first-person pronoun presupposes minimal self-awareness. Thus, if a self-concept is required for minimal self-awareness, one will again be forced to hold that this self-concept is innate, or at least develops very early. Some would no doubt find the idea of an innate (or practically innate) self-concept unproblematic. But such a self-concept would have to be so rudimentary that it becomes questionable whether it counts as a concept at all. In any case, it does not seem to me any *less* plausible to say that infants, say, have various capacities for non-conceptual self-representation.[33]

One might also argue that phenomenal self-awareness requires the subject to possess what we might term 'experiential concepts': concepts that allow the subject to recognize and reidentify certain features of one's experience (as distinct from objects

[33] Of course, one could reject the connection between phenomenal experience and self-awareness. But then one will have to answer the arguments in the previous chapter.

of experience). One type of experiential concept is a phenomenal concept. A phenomenal concept is a concept of a certain phenomenal feature of experience, such as *how red looks, how coffee tastes,* or *how it feels to be angry.*[34] Further, most advocates of phenomenal concepts require that one cannot have a phenomenal concept without having had an experience with the right phenomenal properties. For instance, I can only have the phenomenal concept of how coffee tastes, if I have tasted coffee (or something that tastes just like coffee). On this view, then, I cannot be aware of features of my own experience unless I have experiential concepts. Thus, even if my perceptual states convey information about their own features along with conveying information about the environment, without a repertoire of experiential concepts, I will be P-blind; I will be incapable of awareness of what my experience is like.

Again, the conceptualist position seems too strong. Couldn't a subject have the capacity to recognize features of her own experience without yet possessing full-blown experiential *concepts*? Admittedly, it seems right to say that reflective judgments about the phenomenal character of experience require phenomenal concepts. However, as I have been at pains to show, we have a more basic, prereflective acquaintance with our own experience. Indeed, it also seems right to say that it is *because* we have a prereflective acquaintance with the phenomenal character of our experience that we are able to form the phenomenal concepts used in reflection. That is, in order for me to form the phenomenal concept of how coffee tastes, I need to have a prereflective acquaintance with what it is like to taste coffee; and being acquainted with what it is like to taste coffee involves information about both the coffee and about my tasting of it.

[34] These are very basic phenomenal concepts, but one can of course develop more sophisticated concepts of this type.

The point here is that sensitivity to certain sorts of information, such as information about one's own mental states, is not necessarily a conceptual capacity. And much the same thing can be said about other forms of self-specifying information, such as is involved in bodily self-awareness. I must be sensitive to information about my own body in order to have bodily self-awareness, but it does not follow that this sensitivity requires bodily concepts. An infant may be aware that there is a pain in her left foot without possessing the concepts *pain, left,* and *foot.* What she does require is a body schema, a general representation of her own body. But again, I see no reason to think that having a body schema requires concepts either.

Thus it seems that the types of information involved in prereflective self-awareness have non-conceptual content, insofar as being in such informational states does not *entail* that the subject possess the concepts required to specify the content of those states. Further, I have suggested that the development of the concepts used to specify the content of those states is based on the subject's prereflective sensitivity to the appropriate information. That is, a subject develops the concept *red* in part based on her prior ability to non-conceptually, visually represent objects as red. And she develops phenomenal and bodily concepts in part based on her non-conceptual, prereflective self-awareness (phenomenal and bodily respectively).

However, the conceptualist can challenge this account in two ways. First, one of the key intuitions behind non-conceptualism is that the contents of experience can outrun our conceptual repertoire—e.g. I can perceive more colors than I have concepts for. Yet, John McDowell has argued that this isn't so.[35] According to McDowell, non-conceptualists (he has in mind Gareth Evans) have overlooked the importance of *demonstrative* concepts. I may not have a non-demonstrative concept for every shade I can perceive, but I could have a demonstrative concept—i.e., *that*

shade—for each shade I can discriminate. Hence, according to McDowell, it is not the case that the contents of experience can outrun our conceptual repertoire.

Even if McDowell is right about this, does it follow that perceptual contents are conceptual contents? If sound, McDowell's argument only shows that, thanks to demonstrative concepts, all perceptual experience is *conceptualizable.* Using demonstrative concepts such as *that shade, that feeling,* or *that taste,* may allow us to bring perceptual contents under concepts—and thereby allow us to form beliefs and judgments about them—but it does not follow from this that perceptual contents are conceptual. Moreover, there are independent reasons to reject the idea that perceptual content is conceptual content. As I argued above, perceptual representation is unlike paradigmatic forms of conceptual content in important ways: it is not revisable[36], its contents are not specified in terms of rational connections such as inference, its content is not holistic, etc. Furthermore, it is not yet clear to me that even demonstrative concepts fully capture perceptual content. It seems possible that I could discriminate a shade, but be unable to reidentify that shade at a later time. But surely the capacity to reidentify is required for even the weakest form of concept-possession.

The key point of conceptualism about perception (and other forms of content) is that the concepts possessed by the subject determine what she can perceive; they fix the 'perceptual space' of the subject. The reason for this is that, on the conceptualist view, perceptual representations are made up of concepts. However, the problem with McDowell's attempt to save conceptualism is that he gives up the key point of conceptualism in the process of saving it. This is because once we include demonstrative concepts in the conceptual repertoire, the range of possible perceptions becomes unlimited. Hence, we lose the point of saying that the conceptual repertoire

---

[35] McDowell (1994),

[36] Or at least, not revisable in the way beliefs are. Even when I know I'm seeing an optical illusion, I cannot usually change my perception of the illusion. However, one might argue that gestalt switches

determines the perceptual space of the subject. And besides, perceptual space is *not* unlimited; it is biologically determined.

Now, a second conceptualist response might be that the non-conceptualist is only forced to deny that perception is conceptual because she has an implausibly strict standard for concept possession. This is an important and plausible objection, since it is not uncommon for philosophers to deny that infralinguals can have *any* concepts. I reject that stringent claim. But just because infralinguals *can* have concepts, it doesn't follow that they *must* have concepts in order to have representations. In order to make the claim that concepts are required for representation, one would either have to weaken the criteria for concept-possession so far that it would begin to lose its explanatory power, or else strengthen the criteria for representation so much that one would come into conflict with empirical research on the cognitive capacities of infralinguals. We should try to avoid 'lowering the bar' on concept possession so far that just any recognitional capacity will count as a concept. For instance, a concept should be a relatively context-independent recognitional capacity, one that can be exercised in a wide range of situations that may differ greatly in many respects. In contrast, infralinguals may employ recognitional capacities that are far too situation-dependent to qualify as full-blown concepts, as seems to be the case in perceptual representation. If we loosen the criteria for concept-possession so as to include all (or most) recognitional capacities, we lose our grip on the explanatory power of concept-attribution. Furthermore, even if one does weaken the standards for concept-possession, one still needs to explain why perceptual representation is so unlike paradigmatic forms of conceptual content such as thought and belief.

---

may count as a form of perceptual 'revision'. In any case, even if perceptual states are not revisable, it does not mean that they cannot misrepresent their objects.

Having defended the plausibility of non-conceptualism about perception and prereflective self-awareness, I want to reiterate two important points. First, non-conceptual awareness is *not* 'raw sensation'. Much of the attraction of conceptualism—it seems to me— comes from a picture of the mind divided into indeterminate, non-intentional, raw sensation on the one hand, and determinate, intentional, conceptual contents on the other. On this view, if perception is to have any content at all it has to be conceptual. Non-conceptualism rejects this picture and defends the existence of non-conceptual representation. Non-conceptual representation, like all representation, must have structure and correctness-conditions.[37] Second, non-conceptualism does not entail that a subject's conceptual repertoire has no influence on perception. For concept-users like us, perceptual experience (broadly construed) is a complex affair involving both conceptual and non-conceptual capacities. What I can perceive is affected by, but not fully determined by, my concepts.

4.4 Conclusion

The account of bodily self-awareness that emerges from the above considerations emphasizes that conscious creatures are always embedded in, always interacting with, an environment. It is thus part of a 'bottom-up' account of self-awareness; it starts from the fundamental, on-going active and cognitive interaction of an organism with its environment. As we have seen, the body plays a crucial role in that it is, of course, a condition of the possibility of interaction with the environment. Further, the body (and embodiment) structures experience at practically every level. Moreover, I have argued that bodily self-awareness is—for many creatures at least—

[37] This picture of the mind figured centrally in the thought of both Nyāya and Buddhist philosophers. For instance, for the Buddhists there must be a non-conceptual, unstructured (*nirvikalpaka*) form of perception that provides the foundation for structured, determinate (savikalpaka) perception. My

built into the basic structure of experience. Indeed, rather than being the product of sophisticated conceptual and linguistic capacities, my 'bottom-up' account of bodily self-awareness suggests that it is a non-conceptual form of awareness. We now move to the difficult topic of the self.

---

version of non-conceptualism rejects this dichotomy. Although perception in non-conceptual, it is not unstructured or indeterminate, and thus is not *nirvikalpaka*.

CHAPTER 5
SELF-AWARENESS AND THE SELF

Can one make sense of self-awareness without an appeal to the self? A quick and too sanguine response would be that the term '*self*-awareness' no more presupposes the existence of a self than do the terms '*self*-organizing' or '*self*-regulating'. Yet the types of self-awareness discussed in this dissertation are tightly intertwined with our sense of self—our deep sense that we *have* (or that we *are*) a self. We have in previous chapters explored the particular ways in which we are aware of our own subjectivity, of the way in which we are aware of our experiences *as our own*, as well as our awareness of aspects of our own experiences. To whom (or what) could experiences belong, if not to a self? Who is it that is aware of the phenomenal feel of experience, if not a self? Of whom am I aware when I am self-aware, if not a self? To whom do I refer when I employ the first-person pronoun, if not to my *self*? These are questions not so easily dodged by one who is skeptical of the existence of the self.

What's more, some philosophers—ancient and contemporary—have argued that a proper understanding of first-person contents in fact blocks the kinds of reductionist strategies that are skeptical of our traditional notions of selves (and persons).[1] For instance, John Campbell has argued that the irreducibility of first-person contents to third-person contents—and hence the indispensability of first-person contents in characterizing human mental life—blocks the possibility of a reductionist stance on persons because, the claim goes, an adequate description of reality would have to include some 'personal facts'—i.e., facts about our experiences and their irreducible first-person contents—even if it cannot be shown that persons (and/or selves) are independent, separately existing entities.[2] Similarly, the great Nyāya philosopher Uddyotakara

[1] Reductionism about persons entails either reductionism or outright eliminativism about the self.
[2] Campbell (1994).

questions the compatibility between a correct account of 'I-cognitions' and the Buddhist theory of no-self.

This line of argument is important for my account of self-awareness for several reasons. First, because I have defended the irreducibility of first-person contents, it would seem that, if Campbell is right, then any form of reductionism about persons would be incompatible with my account of self-awareness. While I have so far not relied on any particular account of the self in developing my theory of self-awareness, I may nonetheless be committed to non-reductionism. Second, I happen to be skeptical about the existence of the self and sympathetic towards reductionist accounts of persons such as those defended by many Buddhists and by contemporary philosophers such as Derek Parfit. Hence I need to be able to counter arguments such as those raised by Campbell and Uddyotakara. Finally, and third, certain Buddhist thinkers have defended the reflexivity of awareness (*svasaṃvedana*), and since all Buddhists accept the doctrine of no-self (*anātman*), those who accept both doctrines might also run afoul of these arguments.

Thus, in this chapter, I want to show that my account of self-awareness is compatible with reductionist view of persons, and hence with a rejection of the existence of an substantial, separate self. (Of course, my account is compatible with various non-reductionist accounts of the self as well.) In section 5.1, I will examine reductionism as it relates to accounts of the self. I will then, in section 5.2, argue that a reductionist account of persons can account for the unique features of first-person contents and our deep and multi-layered sense of self. Before I begin, however, I should make clear that my aim here is not to prove the *truth* of reductionism (though I believe that it is likely to be true), but only to show that the view is compatible with my account of self-awareness.

5. 1 Reductionism and No-Self

Reductionism, in this context, is a thesis about the existence and identity of persons. According to Parfit, the reductionist holds that "our existence consists in the existence of a body, and the occurrence of various interrelated mental processes and events. Our identity over time consists in physical and/or psychological continuity."[3] Moreover, on the reductionist view, our existence and identity over time do *not* involve the existence of an ultimate and simple substance, such as a Cartesian Ego or *ātman*. Now, on some views, the self just is something like an *ātman* or Cartesian Ego, and thus reductionism about persons will entail a rejection of the existence of the self. On this view of the self it makes sense to say that there could be selfless persons. However, on a broader view of the self, the notion is often taken to be interchangeable with the notion of a person. In this case, a reductionist about persons could admit that there are selves, but give a reductionist account of them.

In Parfit's characterization of reductionism, it is an ontological thesis, a thesis about the existence and identity of a certain sort of entity. Thus the view should be distinguished from *analytical* reductionism. According to analytical reductionism, each statement in a discourse A—e.g. discourse about persons—has an analytical equivalent in some other discourse B—e.g. discourse about bodies and interrelated mental processes and events—and B does not share any vocabulary with A, other than topic-neutral expressions.[4] Further, discourse B is a philosophically or scientifically 'favored' discourse in comparison to discourse A. A statement in A is analytically equivalent to a statement in B if and only if, a priori and necessarily, the A-statement and the B-statement have the same truth-conditions.

[3] Parfit (1999), p.218 .
[4] Johnston (1997), p.151 .

In contrast, *ontological* reductionism is not a thesis about the redundancy or analytic equivalence of discourses or statements within discourses. Rather, as Mark Johnston notes:

> It is a thesis to the effect that making statements in the discourse in question carries no commitment to entities other than those spoken of in some other, philosophically favoured discourse. Constructions out of the entities described in the favoured discourse may be allowed.[5]

Hence, one can be a reductionist about mental states in that one holds that all mental states are in fact physical states, without holding that the discourse of mental states is analytically reducible to the discourse of physics. One might also be an ontological reductionist about objects such as statues. A marble statue, one might argue, is wholly constituted by the marble that makes it up. Thus the reductionist would say that the statue is nothing *over and above* the hunk of marble and that the existence of the statue *just consists in* the existence of a hunk of marble with a certain shape.[6] Further, a reductionist could argue that the existence of a certain composite entity just consists in the existence of more basic entities arranged in the right way. On this view, the existence of a stereo, for instance, just consists in the existence of appropriately arranged stereo parts. The stereo is nothing over and above these appropriately arranged parts in that it is not the type of thing that could exist separately from the parts that constitute it at any given time.[7]

Reductionism as Parfit uses the term is not only a thesis about the constitution of persons, however; it is also a thesis about the diachronic identity of persons.[8] Our identity over time consists in the holding of certain forms of physical and/or psychological continuity and connectedness. Thus a reductionist might hold that some future person is

[5] Johnston (1997), p.151 .
[6] There are very complicated issues about the status of artifacts, which I will gloss over here.
[7] However, a composite entity could exist independently of the parts that composed it at some previous time. For instance, through replacing the parts of a stereo one by one, the *same* stereo can come to have a completely new set of parts.
[8] Note that Reductionism is compatible with the view that we are not essentially persons. So for instance a Reductionist might hold (as Parfit now does) that we are essentially human beings and that *person* is a phased-sortal like *adult* or *adolescent.*

the same as some present person just in case they are uniquely psychologically continuous. On the other hand, some reductionists hold that it is physical continuity, such as the continuity of a body, that is fundamental to our identity over time. In addition, on Parfit's own version of reductionism, our identity need not be determinate in all possible cases. "We can," he argues,

> imagine cases in which questions about our identity would be indeterminate: having no answers. These questions would also be in the following sense *empty*: they would not be about different possibilities, but only about different descriptions of the same course of events. Even without answering such questions, we could know what would happen.[9]

The idea here is that since the facts about our diachronic identity just consist in the holding of more basic facts about physical and/or psychological continuity and these more basic facts can be a matter of degree, our identity can be indeterminate. Furthermore, Parfit argues that since questions about our diachronic identity will not always have a 'yes or no' answer, our existence and diachronic identity cannot involve a Cartesian Ego, or other such entity. Not all reductionists, however, hold that our identity could be indeterminate.

5.1.1 Buddhist Reductionism

Having sketched the basics of general reductionism about persons, let me now turn to the Buddhist *anātman* theory. With the exception of the Pudgalavādins, the doctrine of *anātman* (no-self) is given a reductionist interpretation in Indian Buddhist philosophy—an interpretation in keeping with the more general reductionist orientation of much Buddhist ontology. While this section will be concerned with the reductionist understanding of the doctrine of no-self, we may begin with what unites both reductionist and non-reductionist interpretations of the doctrine: the rejection of a substantial, separately existing, permanent and unchanging self. First and foremost, the doctrine of

no-self is a rejection of the *ātman*—the unitary, unchanging, eternal, spiritual substance that is said, in the Vedic tradition, to be one's true self. And this rejection of the *ātman* is closely tied to the Buddhists' reductionism about persons (*pudgalā*).

On the Buddhist view the existence of a person just consists in the existence the five *skandhas* arranged in the right way. The five *skandhas* are the body or physical form (*rūpa*), feelings (*vedanā*), perceptions (*saṃjñā*), dispositions (*saṃskāra*), and consciousness (*vijñāna*). A person, on this view, is not an entity that can exist independently of the five *skandhas*—a *pudgala* is, as the reductionists say, nothing 'over and above' the five *skandhas* appropriately organized. As a generic term, we may call an appropriately organized collection of the *skandhas* a 'psychophysical complex'. The Buddhists argue that none of the *skandhas* individually, nor the complex of *skandhas* collectively could be the self—i.e., the independent, substantial, unchanging, inner controller and owner of the *skandhas.* Take away the complex, impermanent, changing *skandhas* and we are not left with a constant, substantial self; we are left with nothing.

Furthermore, the diachronic identity of a person just consists in the appropriate degree of continuity and connectedness of the *skandhas.* The various mental and physical factors that constitute the person are not each connected to an enduring self, but rather they are only connected to one another. And since the *skandhas* include both psychological and physical factors, the standard account of our identity will be a matter of both physical and psychological identity. However, since the Buddhist account of rebirth does not involve physical continuity, psychological continuity is, on their view, sufficient to account for identity across lives. Thus, the synchronic identity of a person is a matter of there being an appropriately organized complex or bundle of *skandhas*, whereas the diachronic identity of a person is a matter of there being a causally and functionally integrated series or stream of *skandhas* (*skandhasaṃtāna*).

---

[9] Parfit (1999), p.218 .

The Buddhist account of the person is part of their general reductionist and anti-substantialist ontology.[10] That is, they are reductionists about *all* composite entities and they reject the substantialist theory of objects in favor of a bundle theory. On the Buddhist view, a composite entity is nothing over and above a certain arrangement of its parts. That is, the whole cannot exist apart from its parts (it is dependent on its parts) and its intrinsic properties are derived from the properties of the parts and how those parts are related. Moreover, wholes are causally epiphenomenal in that what appear to be the causal powers of the whole can in fact be accounted for in terms of the causal powers and interactions of its parts.[11] So, on the Buddhist view, composites are constructions of their basic parts.

We can get a clearer idea of this form of compositional reductionism by briefly considering one of the earliest formulations of it in the Buddhist canon. In the *Milindapañha*[12], the Buddhist monk, Nāgasena, debates the Greco-Bactrian king Milinda on a variety of topics central to Buddhism. In the famous first chapter, a person is compared to a chariot. In debate with Milinda, Nāgasena shows that a chariot is not any one of its parts, it is not the mere sum of its parts, and yet it is not anything separate[13] from its parts.

What, then, is the chariot? The answer finally arrived at involves a 'semantic ascent' to the application of the term 'chariot'. "It is depending on the axle, the wheels, the chassis, and the pole, that there is this designation, appellation, conceptual construct, convention, mere name, i.e., 'chariot'."[14] The idea here is that we apply the term 'chariot'

[10] Unless otherwise noted I will the term Buddhist to refer to 'mainstream Buddhism'—i.e., primarily early Buddhism, Abhidharma, and the Buddhist Logicians.
[11] So, like such philosophers as Jaegwon Kim, Buddhist reductionists would deny the existence of emergent causal powers.
[12] Kapstein (2001), .
[13] Reductionists such as Parfit hold that while wholes are not *separate* from their parts (because they cannot exist independent of their parts) they are nonetheless *distinct* from their parts (because they have different persistence conditions from mere sums of their parts). Other reductionists hold that constitution is identity. The point in the *Milindapañha* is that there is no independent and unchanging thing underlying the complex and changing chariot.
[14] Kapstein (2001), p.90 .

to an appropriately organized collection of chariot parts. When the various parts come together in the right way, we label the resulting complex a chariot, but the chariot is not an entity over and above this complex. Yet, on the Buddhist view, we have an almost irresistible tendency to reify composite entities (especially ourselves), taking them to be irreducible, substantial, and unchanging. Hence we should understand the Buddhist emphasis on the reducibility of composite entities as an attempt to counteract this tendency, a tendency that, on their view, is in large part responsible for the frustration and suffering (*duḥkha*) that characterizes the human condition.

Recall that, on Parfit's reductionism, certain questions concerning reducible entities will turn out to be empty. So, for instance, certain questions about the diachronic identity of persons will have no answers, even though we have all the facts about the integrated series of mental and physical events and processes in which our identity consists. For Parfit, our answers (or lack thereof) to certain questions about personal identity will depend more on how we choose to describe a course of events than on the nature of that course of events. Interestingly, Buddhist reductionists seem to hold a similar view. For them, our common sense ontology for composite objects is partly conventional. We choose to label certain complexes but not others based on our conventions, needs, desires, and karmic tendencies. This is the point of Nāgasena's semantic ascent, his emphasis on the fact that *we* use terms such as 'chariot' and 'Nāgasena' to pick out complexes that are important to us. And again, this emphasis is, it seems to me, in the service of counteracting the human tendency to reification.

Unlike many reductionists in the western tradition, however, the Buddhists explicitly endorse and rely on a two-tiered ontology. Reducible entities are conventionally real (*saṃvṛtisat*), whereas basic, irreducible entities are ultimately real (*paramārthasat*). All conventionally real entities are constructions out of ultimately real entities. And, of course, since all real composite entities are *conventionally* real on the

Buddhist view, ultimately real entities must be simple. Corresponding to this ontological distinction, the Buddhists also have a two-tiered conception of truth. Conventional truths are those truths that quantify over conventionally real entities, whereas ultimate truths only quantify over ultimately real entities.[15] Further, the discourse of ultimate truth is the Buddhists' 'philosophically favored discourse'. When we use conventional discourse, we are not ontologically committed to anything but the entities mentioned in the ultimate discourse (including constructions out of such entities), even if conventional discourse is not analytically reducible to ultimate discourse. Moreover, conflation of the conventional and the ultimate is central to the Buddhist diagnosis of the human tendency toward reification. For instance, persons are conventionally real, but we have a strong tendency to understand persons as irreducible, substantial, ultimately real entities.

The basic, ultimately real entities of the Buddhist ontology are called *dharmas*. All conventionally real entities are constructions of *dharmas*. *Dharmas* basically fall into two categories: mental and physical. So, for example, material objects are made up of some combination of earth, air, fire, and water *dharmas*. However, it is important to note that material *dharmas* are not little bits of material *stuff*. Rather, *dharmas* are particular properties, or what contemporary analytic metaphysicians would call *tropes*. Thus, for example, an earth *dharma* is a particular instance of solidity and a water *dharma* is a particular instance of fluidity. In addition, *dharmas* are causally efficacious events or occurrences.[16] Finally, *dharmas* are momentary (*kṣanika*), lasting only long enough to be causally efficacious.

On the substantialist account of objects, objects are substances in which properties inhere. Further, on the substantialist account of change, one and the same substance endures through time, gaining and losing properties. But on the Buddhist view, objects

[15] While the Buddhists did not characterize the distinction in terms of quantification, their criterion is based on existence and identity. Hence the rephrasing in terms of quantification is apt.

are simply bundles of *dharmas* and the *dharmas* do not inhere in a substance at all. Moreover, the Buddhists obviously cannot appeal to enduring substances to account for identity through change. Rather, change is accounted for in terms of causally integrated series of momentary *dharmas*.[17]

But in order to unpack this account, let me distinguish between persistence, endurance, and perdurance. I will use 'persistence' as a generic term that applies to any object that exists for more than one moment. Hence, there are two main accounts of persistence. On the one hand, substantialists account for persistence in terms of enduring substances. An object *endures* if it is wholly present for more than one moment. On the other hand, the Buddhists can be seen as advocating a perdurance account of persistence. An object *perdures* if it has temporal parts at more than one moment. Such an object (a continuant) will not be fully present at any given moment, since it will also have at least one temporal part in either the present or the past or both. For instance, a theatrical performance has a beginning, middle, and end, and so is not fully present at any given moment during the performance.

As anti-substantialists, the Buddhists deny the existence of enduring substances. In addition, as we have seen, they hold that *dharmas* are momentary. But conventional objects such as chariots and people do persist.[18] How is this possible? The persistence of an object is explained in terms of the existence of an integrated series of momentary stages. No basic part of the series lasts more than a moment, but the *series* of parts persists in virtue of having temporal parts at more than one time.[19] So, for instance, a pot

[16] Most contemporary trope theorists would agree, since they hold that an event is a particular instantiation of a trope or complex of tropes. Others hold that events are themselves tropes.
[17] Change, for the Buddhists, must be accounted for as the arising and passing away of momentary *dharmas*.
[18] The Buddhists see their view as a middle way between the extremes of eternalism (belief in unchanging enduring substances) and annihilationism (belief that nothing persists).
[19] To the extent that *dharmas* are events (or event-like) we can see the Buddhist ontology as an early instance of process metaphysics. According to Rescher: "A process is a coordinated group of changes in the complexion of reality, an organized family of occurrences that are systematically linked to one another either causally or functionally." Rescher (1996), p.38 . This nicely fits the Buddhist view, which is thus a form of process atomism.

is taken to be a series of connected pot-stages (the stages are themselves bundles of *dharmas*). Despite the fact that there is no enduring substance, we can identify (roughly) when the pot came into and went out of existence.

Moreover, changes in the pot will be accounted for in terms of the differences in the *dharmas* making up different stages of the series. Thus, if I make a pot on Tuesday and paint it red on Thursday, we can say that this is a change in the *same* pot because the later red pot-stage is appropriately connected to the earlier pot-stages; they are parts of the same continuum (*saṃtāna*). The identity of the conventional object, then, is a matter of the continuity of the momentary *dharma*-bundles that make it up. And the same is true for persons. Two person-stages are part of the same person just in case they are part of an appropriately continuous and connected series of physical and/or psychological events (*dharmas*) and processes (integrated sub-series of *dharmas*). As Nāgasena points out to king Milinda, the identity of a person over time is like a candle that burns all night. The flickering flame is just a series of momentary events, but these events are causally connected in such a way that it is appropriate to say that the same flame burns all night. In the same way, I am the same person as Matt at five years old, not because we share an enduring self, but because that child is uniquely continuous with me.

Furthermore, on the Buddhist view, our common sense ontology (and languages) is substantialist and non-reductionist. We interact with *dharmas* and series of *dharmas*, but through a process of conceptualization, we come to see *dharma*-constructions as enduring substances. But since there are no such enduring substances on the Buddhist view, conventionally real objects *as we normally conceive of them* do not exist. On my view, this explains why there are many passages in Buddhist texts that seem to suggest that conventionally 'real' objects are completely unreal. Indeed, such passages have led some contemporary interpreters of Buddhism to posit that Buddhists are actually mereological nihilists—that is, that the Buddhists hold that in fact no composition ever

occurs. On this view, all conventionally real entities are unreal, though we talk as if they exist, and all conventional truths are actually false, though we treat them as true in everyday contexts. However, this view does not account for the fact that many Buddhist texts and philosophers seem quite happy to admit bundles, series, and other composite objects into their ontology, so long as it is understood that they are reducible to arrangements of *dharmas.* In any case, as I take Buddhist reductionism, it is reified wholes and enduring substances that are its prime targets, not composition *per se.* On this view, there are no chariots or people as we normally conceptualize them—that is, as enduring substances—but the Buddhists can give an alternative account of the objects of experience that reflects the three marks of existence—impermanence[20], *duḥkha*, and selflessness[21]—in their metaphysics.

5.2 Self-Awareness without a Self

As I mentioned above, my aim here is not to prove that reductionism is the correct account of persons, though I think that some form of reductionism is probably true. I have the narrower aim of showing that one can be both a reductionist and a reflexivist. Yet even this narrower aim is faced with several powerful objections, from non-reductionists in both the Indian and western traditions. In this section I will address these objections and then sketch an account of our sense of self that is compatible with a rejection of the existence of the self.

5.2.1 The First-Person

If a particular token of the first-person pronoun 'I' does not refer to the utterer's self, what, if anything, does it refer to? It seems glaringly obvious that when an individual

[20] On the above interpretation impermanence is taken to mean that nothing *endures* and that the world is marked by constant change, since momentary *dharmas* are constantly coming into an going out of existence. Thus impermanence is not incompatible with *persistence.*

correctly uses 'I', she is referring to *herself.* But from this it does not follow that when she uses 'I' she refers to her *self.* Self-reference is not necessarily reference to a self. So even if one rejects the existence of the self, one can still give an account of first-person self-reference.

The reductionist could take a Wittgensteinian route and claim that 'I' does not refer at all. Thus, the problem of what 'I' refers to is circumvented, though other problems with this approach may arise. The reductionist might also endorse the more commonsensical claim that 'I' refers to the person who utters it, so long as we understand persons in the reductionist way. It is no part of reductionism that terms for reducible entities fail to refer. And since reductionists hold that persons exist, I see no reason to say that 'I' does not refer to them. What the reductionist denies is that persons *as conceived by non-reductionists* (hereafter, Persons) exist. Further this option would be especially attractive to reductionists, such as Parfit, who hold that constitution is not identity—that is, that a person is constituted by a body and series of interrelated mental events and processes, but is not identical to this psychophysical continuum. A third option is to say, as the Buddhists do, that 'I' refers to a psychophysical continuum. On this view, the reductionists gloss the rule for the use of the first-person as, "'I' refers to the psychophysical continuum that produced it." Now, of course, we may not *intend* to refer to a psychophysical continuum when we use 'I'. However, since the reference of 'I' is not fixed by definite descriptions, there can be a gap between our intentions and the actual reference of 'I'.[22] Perhaps we all think we are referring to selves or Persons when we use the first-person, but in fact, the reductionist says, we are just referring to psychophysical

[21] The term here is *anātman*, but it is applied to all things. all things are selfless in that there are no substances.
[22] Since 'I' is an indexical, its reference is context-sensitive, but also rigid. Thus even if I believe that I am the Queen of England, when I say "I am the Queen of England" it still refers to Matt MacKenzie and not the Queen.

continua.[23] However, it seems to me that the second and third options are not all that different. I see no reason why the Buddhist cannot admit that 'I' refers to persons, since on their view persons *just are* psychophysical continuants.[24]

At this point it might be objected that *whatever* 'I' refers to is the self, thus the reductionists are wrong to deny the existence of the self. But this objection is no help to the non-reductionist. If that is all the self is, then one cannot appeal to the self as that which individuates persons (or psychophysical continua), nor can one say the self is an enduring thing. The self will no longer do any important work.

A more sophisticated objection arises from the Nyāya philosopher, Uddyotakara. He writes:

> The consciousness of "I," which conforms to the distinctions of the nature of the object, and which does not depend upon memory of marks, the possessor of the marks, and their relationship, is direct acquaintance just as is the cognition of physical form. Concerning what you yourself, with perfect confidence, establish to be direct acquaintance, in virtue of what is it that it is [said to be] direct acquaintance? You must establish it as being consciousness alone, which does not depend upon the relationships among marks, etc., and which is self-presenting. So then you think there is an I-cognition, but that its object is not the self? Well, then show us its object![25]

As we saw in Chapter 2, and as Uddyotakara points out here, first-person self-reference must be anchored in a non-criterial, non-inferential mode of self-acquaintance. Moreover, if, as the reflexivists argue, I-cognitions are self-presenting, the awareness of oneself as a subject is built into the structure of I-cognitions. But if there is no self, what are I-cognitions directly acquainted with? What is the subject of experience?

---

[23] This is because the use of 'I' is relatively *ontologically neutral*—that is, it is compatible with both non-reductionist and reductionist theories of the person. Thus we could imagine a society of reductionists who use 'I' correctly in their linguistic practices even though reductionism is false. Hence my view is that we may think 'I' refers to a substantial self, but it does not. This is in contrast to the view that 'I' *does* refer to substantial selves, but that such entities do not exist.

[24] There are good historical reasons for wanting to avoid talking about 'I' as referring to persons, though. Since the Pudgalavādins attempted to defend a non-reductionist account of persons while rejecting the *ātman*, mainstream Buddhists would not have wanted sound like this quasi-heretical school.

[25] Kapstein (2001), p.98 .

The Yogācāra answer to this challenge is that, as Uddyotakara says, it is consciousness alone that is the subject of experience. Of course, for the Yogācārins, there could be no other answer, since they hold that consciousness is the only thing that exists anyway. (But Sartre, who is not an idealist, would give a similar answer.) Further, the Yogācārins actually hold that the subject of an experience just is the subjective aspect (*grāhakākāra*) of that very experience. And since experiences are momentary the subject of experience does not persist. Thus the 'I' in each of two successive I-cognitions in fact refer to different subjects.

There are serious problems with this view. For one thing, it seems to needlessly multiply subjects. On this view a single life involves a huge number of different subjects. Moreover, if 'I' refers to part of a momentary stage in a mental continuum, then it follows that 'I' does not refer to the same entity that began writing this sentence. But, as we saw in the *Milindapañha*, most Buddhists would say that 'I' does refer to the same entity across times, even if they reject the substantialist account of why this is the case. However, perhaps it is open to someone who holds the view in question to respond that we should not take 'the same' to mean 'numerically identical' in this case. Each stage of a mental continuum has a subject and the subject of a stage at time t is the same as the subject of a stage at time $t+1$ just in case they are appropriately related. Thus 'the same' in this case will track, not numerical identity, but some weaker unity relation such as causal or psychological continuity. Further, the defender of this view could argue that when counting diachronically, we count appropriately related sums of stages, not individual stages. Hence, we can say that typically there is one subject per life.

I am not convinced that these responses are adequate. However, the reductionist does not have to take the Yogācāra view. Rather, she can say that the subject of experience just is the continuum as a whole. This is Vasubandhu's view.[26] 'I' refers to the

[26] That is, his view before he became a Yogācārin.

psychophysical continuum (*skandhasaṃtāna*), not a particular stage of that continuum. Moreover, the subject of experience, the agent of actions, is not an entity distinct from the continuum—i.e., it is not a *self*—it is the continuum itself. It follows from this view that the subject is distinct from any given experience, since the series is not identical to any one of its parts, but it is not a separately existing thing over and above the stream of experience. And since the subject is a continuant, it cannot be an enduring substance such as a Cartesian Ego or *ātman*.

But Uddyotakara could now object that the reductionist has missed his point. He was arguing that the subject is directly presented to itself in I-cognition. And this seems to imply that self is an object directly available to the I-cognition (and therefore available to itself) at the time of the cognition. Yet if the subject is a temporal series, then *ex hypothesi*, the subject is not wholly present at any given moment. At best the reductionist can say that only part of the subject is revealed in I-cognitions, and part is distinct from whole. Therefore, the subject isn't really revealed.

This argument fails, though, because we *can* be aware of an entity by being aware of one of its parts. For instance, I am presently visually aware of my computer because I see its front. And more to the point, when I listen to Coltrane's "A Love Supreme" I am hearing the song in virtue of hearing its beginning, middle, and end in succession. At each point, I am listening to a song, even though the song is never wholly present. And when I cross a river, it is the river I am crossing, not merely a stretch of it. Thus it seems we can be aware of ourselves as subjects even if subjects are continuants, not substances.

Finally, it has been argued that a proper understanding of first-person reference blocks a reductionist account of persons. Parfit initially argued that reductionism about persons entailed that we could give a complete description of our lives without referring to ourselves, or explicitly claiming that we exist. If the first-person is irreducible, however, then we will be unable to give a complete description of our lives in impersonal

terms. Further, if the content of an indexical thought, such as an I-thought, depends in part on what that thought is about, then we cannot describe this thought without claiming that it is about a person.[27] And this too conflicts with impersonal redescription of ourselves.

In light of these objections, Parfit has recently abandoned the impersonalism criterion for reductionism. He now holds merely that "when experiences at different times are *copersonal*—or had by the same person—this fact consists in certain other facts." And further that "these other facts must be describable in a way that does not assume the copersonality of these experiences." Thus, for instance, if copersonality consists in various sorts of causal and functional interrelations, then the reductionist can describe these interrelations without relying on copersonality to explain them. If I want to explain why it is that several components are parts of the same computer, I might explain that part A is connected to part B in such a way that C can function properly in relation to D, and A, B, C, and D are encased in a plastic shell, such that one could grab the shell and move all the parts from one location to another. And when all the parts are related in just the right way, the computer works. On this view, it is appropriate to assume that the computer exists, to refer to it, and to rely on our concept of a computer. What is not appropriate is to say that A is part of the same computer as B because they each have the relation of belonging to some computer X. That is, the reductionist cannot treat belonging to the same computer as a brute or ultimate fact that does not consist in the holding of more particular, more basic facts.

Furthermore, the main problem with the objections from the reference of the first-person is that reductionism is an ontological thesis, not an analytical thesis. Thus the reductionist need not hold that 'I' is reducible to some impersonal form of reference. And the fact that the content of an indexical involves the object it is about is compatible with

---

[27] Cassam (1992), .

reductionism. Just because the statement "*This* statue is marble" involves the statue itself does not entail that the statue is not wholly constituted by the statue-shaped hunk of marble that makes it up.

5.2.2 Subjects and Experiences

The reductionist thinks it possible to give an account of the existence and identity of persons in terms of certain more particular, more basic facts. That is, the reductionist believes that one could give a substantive account of what the existence and identity of persons consists in. For such an account to be possible, it must be possible to identify and explain these more basic, particular facts without relying in any substantive way on the concepts such as 'person' and 'copersonality'. But as we have seen above, this does *not* mean that the reductionist cannot in any way utilize these concepts; it only means that she cannot assume these concepts as part of the explanation. It is perfectly legitimate for the reductionist to rely on her concept of a person in giving her explanation in that she is attempting to give an account of persons, and so must have the concept 'person'. Moreover, she will need to keep our general concept of a person in mind as she constructs her explanations in order to see how (or whether) her account of the more particular, basic facts coheres with the general concept of a person.

However, non-reductionists argue that experiences depend on subjects in so intimate a way that one cannot describe the interconnected flow of experience without substantively relying on the notion of a subject. That is, they argue that there is no (relevant) level of facts more basic than those constitutively involving subjects. Thus reductionists cannot explain persons and personal identity in more basic terms; reductionist explanations will always be circular because there is no account of experience that doesn't appeal to subjects in a way that reductionists want to avoid.

There are at least three ways to flesh out this objection. First, relying on a substance metaphysics, it could be argued that events, including experiences, require substances. All events and processes must involve changes in the attributes of a substance or substances. However, taken as a general thesis, this view is hard to maintain. There seem to be many instances of events and processes that do not essentially inhere in a substance or substances: a flash or bang, a hurricane, a musical performance. If we accept that there are substances, we can admit that these events or processes *involve* substances. For instance, a hurricane may involve flying debris, or a musical performance may involve instruments. But it does not follow that we must conceive of hurricanes or performances as belonging to a substance.

Nonetheless, one could argue that while the substance-attribute model does not apply across the board, it *does* apply to experiences. On this view, an experience is not like a state of the weather; rather, an experience essentially belongs to a subject and this is best understood on the substance-attribute model. Yet even if one admits that experiences must belong to a substance, it does not follow that we must reject reductionism. Suppose a reductionist is also a physicalist. She might argue that experiences essentially inhere in a substance, but that the relevant substance is not a subject or self in the appropriate non-reductionist sense. Rather, the relevant substance is the brain. All experiences either are or are realized by states of the brain and in this sense they fit the substance-attribute model. Or, suppose that the reductionist is a dualist. On this view experiences might necessarily inhere is some sort of mental substance. But it does not follow that this mental substance is the self or subject. Rather, it might simply be a kind of impersonal 'mind-stuff'. In either case, the reductionist can appeal to facts about the substance in which experiences inhere without illegitimate reliance on the notion of a subject (or person, etc.). For instance, the physicalist might rely on an account of the existence and persistence-conditions of brains and the conditions under which two

experiences are had by the same brain as part of her reductive account of persons, without appealing to the copersonality of experiences.

A second way in which the objection could be fleshed out is by an appeal to the structure of our discourse about experience. According to Shoemaker, experiences are *experiencings* and since the term 'experiencings' is the gerund of a verb, it implies an *experiencer.* This type of argument has also been used in the Indian tradition, with appeal to Sanskrit grammatical categories. But, as it stands, it is a relatively weak point. Thunderings also imply a thunderer, but it would be dubious to claim that we are therefore committed to the existences of thunderers. However, lurking in the background of this grammatical argument is the notion of conceptual dependence. It could be argued that if the concept of a subject is conceptually prior to the concept of an experience, then reductionists could not give a reductive account of subjects or persons. For example, if the concept 'man' is conceptually prior to the concept 'brother', then I cannot rely on the latter to explain the former. Indeed, even if two notions are interdependent, that would seem to preclude reductive analysis.

Now, it is not clear to me that the concept of a subject is conceptually prior to the concept of experience, though they might be interdependent. But these issues are beside the point, since, as I have pointed out previously, the reductionist is not after a conceptual analysis of the concept 'subject' or 'person'; that is, reductionism is an ontological thesis, not an analytical thesis. Reductionism is based on a claim of ontological dependence—in particular, the notion of compositional dependence. This is important because conceptual and ontological dependence can run in opposite directions. For example, if, as Strawson argues, our concept of a subatomic particle conceptually depends on our concept of mid-sized material objects, it is still the case that these large objects compositionally depend on subatomic particles. Further, the concepts 'trunk' and 'branch' may conceptually depend on the concept 'tree', but it is still the case that trees are constituted by their trunk

and branches. Therefore, an appeal to the conceptual priority of the notion of a subject does not disprove reductionism.

This brings us to the third approach. The non-reductionist could argue that experiences are *adjectival* on subjects in a way that involves both ontological and conceptual dependence. Thus an appeal to the interconnected flow of experiences cannot do the work required of it by reductionism. As an example of adjectival dependence, we might think of the relation between dents and surfaces or the relation between a branch and a branch-bending. As Shoemaker argues, "the ontological status of an experiencing . . . is similar to that of a bending of a branch . . . an experiencing is necessarily an experiencing by a subject of experience, and involves that subject as intimately as a branch-bending involves a branch."[28] According to Parfit:

> For Xs to be *adjectival* on Ys, we might require at least the following:
> (1) Xs are, essentially, *of* or *in* Ys.
> (2) There could not be Xs without a Y.
> (3) An X of one Y could not have been an X of a different Y.[29]

A dent, for instance, is essentially of or in a surface. There could not be a dent without a surface. And a dent of a particular surface could not have been a dent on another surface. In addition, a chess move is essentially a move in a game; there could be no chess moves without chess games; and a particular move in a game could not have been a move in another game (though of course the same *type* of move could occur in another game).[30]

Now, on a Humean account of experiences—and perhaps on certain Buddhist accounts as well—experiences are not adjectival on subjects. They are not essentially of a subject, they are independent, and an experience had by one individual could have been had by another. But a reductionist need not accept this view because, like conceptual and ontological dependence, adjectival and ontological dependence can run in opposite

[28] Shoemaker (1996), p.10 .
[29] Parfit (1999), p.239 .
[30] Parfit (1999), p.239 .

directions. Thus, for example, although chess moves adjectivally depend on games, it is nonetheless the case that chess games are (at least largely) constituted by an appropriately organized[31] series of moves.

Hence, the reductionist can argue that even if experiences are adjectival on subjects, it does not follow that the kind of compositional dependence they advocate is precluded. An experience may essentially belong to a particular subject; it may be that there can be no experiences without subjects; and it may be that a particular experience could not have belonged to a different subject. That is, experiences may be adjectival on the subject (or person) that they together constitute.[32]

Even Buddhist reductionists, who tend toward a more Humean view, could accept this position. As we have seen, Buddhist reductionists hold that persons (*pudgalā*) are composed of the five *skandhas* appropriately organized. Further, the *skandha*-complex is itself composed of momentary mental and physical *dharmas*. It seems to me that it is open to the Buddhist to argue that mental *dharmas* are adjectivally dependent on the psychophysical series in which they occur, while also arguing that psychophysical series are constituted by *dharmas*. Indeed, one could develop such a view starting from the central Buddhist notion of 'dependent origination' (*pratītyasamutpāda*). Mental and physical *dharmas* arise in dependence on a complex of causes and conditions, and in turn serve as the causes of the production of future *dharmas*. Moreover, they are, like tropes, essentially particulars. Thus the Buddhist reductionist could argue, with regard to some particular mental *dharma*, that while the same type of *dharma* could arise in a different *skandhasaṃtāna*, *this* particular *dharma* could only have arisen from *these* particular causes and conditions involved in *this* particular psychophysical series. That is, the

[31] 'Organized' here should be taken to involve a variety of modes of organization. The moves must occur in the right temporal order, but they must also be 'organized' in accordance with the rules of the game.
[32] Parfit (1999), p.239 .

Buddhists could appeal to a form of essentialism about origins.[33] Secondly, if the arising of a certain mental *dharma* requires a specific complex of causes and conditions, there is nothing to stop the Buddhist from arguing that that mental *dharma* could not arise unless there already exists a psychophysical complex of a certain kind. For instance, it could be argued that in order for a linguistically structured thought to arise, there must be a psychophysical complex that is capable of having perceptions, learning language, etc., that is an essential part of the relevant causal complex. Thus there could be no thoughts without a psychophysical complex. Finally, it is open to the Buddhist to argue that certain mental *dharmas* can only exist if they are causally (and perhaps functionally) integrated with other mental *dharmas*. And if such interrelations can only occur in a psychophysical complex, then certain mental *dharmas* are essentially of or in such a complex.

Of course, this is no more than a gesture in the direction of a full account. The point, however, is that one could admit that experiences are adjectival on psychophysical series without thereby committing oneself to selves, subjects, or persons in a non-reductionist sense. Yet, the reductionist still owes an account of the appropriate causal and functional connections that allow for the creation of distinct 'psychological spaces'—that is, what individuates different streams of experience?—an account that does not appeal to copersonality. While I cannot fully address the issue here, I can note that this question is beginning to be addressed by reductionist philosophers. For instance, Bermudez suggests that the appropriate causal and functional connections between first-person mental states could be explained in terms of suitable neural pathways between mental states in the same central nervous system. He goes on to suggest:

> [T]he internal coherence of a psychological space can be captured without a common owner for all the mental states in the psychological space. It is a function of the coherence of content across all the perceptually-based states within that psychological space. By a perceptually-based state I

[33] Note that this form of essentialism does not contradict the anti-essentialism of Buddhists such as Nāgārjuna. Indeed, one way to cash out his linking emptiness to dependent origination is to say that *dharmas* are essentially dependent on other *dharmas* in a way that precludes *dharmas* having *svabhāva*.

> mean one whose content is partly perceptual, and this will obviously include certain types of memory and certain types of intention.

Moreover, the internal coherence of perceptually-based mental states is grounded in the "spatio-temporal continuity of the body" and, as I have argued previously, these states reflect a perspective on the world. Thus we may be able to explain the basis of the unity of psychological space in terms of the continuity of perspective through space and time—a continuity not of an inner self or even a Person, but of a psychophysical system. Furthermore, the fact that systems like us can keep track of their own mental and physical states, I believe, will be crucial to understanding how we construct representations of ourselves as persisting subjects (even if there are no enduring selves).

What is now needed is an account of our very real *sense* of self, an account that is compatible with both reductionism about persons and with the forms of prereflective self-awareness I have been defending.

5.2.3 The Sense of Self

Despite the reductionist rejection of a self, it cannot be denied that we have a deep and multi-layered *sense* of self. We don't just believe we have a self; the sense of self is a fundamental part of our experience. How are we to understand this complex sense of self without positing a self? The beginnings of an answer are to be found in the forms of prereflective self-awareness I have defended.

Like other organisms, human beings are homeostatic systems; we must maintain our integrity by self-regulation. We are, in addition, equipped with the ability to distinguish ourselves from our environment, the 'me' from the 'not-me'. We are also equipped with various more sophisticated modes of self-representation and self-awareness. We are prereflectively aware of our own conscious states as well as certain bodily states. Indeed, as we have seen, neuroscientists have discovered that the brain

utilizes several models of the organism of which it is a part and integrates these models with models of the external world. Further, recall that Sommerhoff argues that the integration of these models constitutes an 'integrated global representation' (IGR), which he identifies with consciousness.[34] Furthermore, consciousness is also reflexive in that it 'illuminates' itself in the act of 'illuminating' its object. Finally, the internal coherence of 'psychological space' and our on-going awareness of our bodily being give rise to a rudimentary sense of the continuity of our perspective through time. Let's look more closely at this picture.

Sommerhoff has argued, on the basis of his neuroscientific research, that, "the brain forms an extensive internal representation of the current state of the organism which includes representations of the total situation facing the organism both in the *outer* and the *inner* world."[35] On Sommerhoff's view—and my own—basic level consciousness is self-referential in that it constitutively involves self-specifying information, such as awareness of one's own bodily and mental states. Furthermore, the IGR is structured into foreground and background by the faculty of attention: that to which one attends forms the foreground of awareness, while that to which one is not presently attending constitutes the background of consciousness. Note, however, that background consciousness is still *consciousness.*

Now one of the most important and stable features of the IGR is the *body schema*, "a coherent structural representation in the brain of the spatial relations involved in body posture, body movement and body surface."[36] In addition, the body schema is essential to the over-all subjective *feel* of the body (from the inside, as it were), or what Merleau-Ponty calls our bodily subjectivity. Again, parts of the body schema, according to Sommerhoff, are incorporated into the IGR, and are thus a part of consciousness.

---

[34] Sommerhoff (2000), .
[35] Sommerhoff (2000), p.12 .
[36] Sommerhoff (2000), p.165 .

Therefore, the body schema manifests itself experientially at even the most basic level of awareness, even though this aspect of our experience is not always (or even often) attended to. Further, compared to the ever-changing representation of the external world, our bodily self-representation is quite stable. What the integration of the body schema into consciousness provides, then, is a stable, continuous background awareness of our own bodily subjectivity, which is integrated with an awareness of an ever-changing external world. As William James has aptly characterized it:

> Our own bodily position, attitude, condition, is one of the things of which some awareness, however inattentive, invariably accompanies the knowledge of whatever else we know. We think; and as we think we feel our bodily selves as the seat of our thinking.

When this continuous sense of bodily subjectivity is combined with a prereflective awareness of the subjective aspects of our experiences—both in that we are aware of the phenomenal quality of all conscious experiences and in that we are implicitly aware of the *perspectivity* of those experiences—we have the basic components of a rudimentary sense of self. But this sense of self, while it is the condition of the possibility of a more robust sense of self, is really only a prereflective awareness of our own *subjectivity*. We need now to discuss the development of a sense of *identity*.

Self-awareness is a many-layered phenomenon, involving, at its most basic, nonconceptual awareness of our own subjectivity, and, at its most advanced, the contemplation of our deepest nature, or of our lives as a whole. Sophisticated forms of self-awareness require sophisticated forms of self-representation, involving a *concept* of oneself and the ability to explicitly ascribe to oneself various psychological and physical predicates. In addition, our sophisticated forms of self-awareness involve an explicit conception of ourselves as existing through time. Once an individual has developed a concept of herself and the temporality of her existence, we can say she has achieved *autobiographical* self-awareness. Beyond autobiographical self-awareness, typical human

beings are able to engage in *contemplative* self-awareness. This most sophisticated form of self-awareness involves the ability to consider what kind of person one is, or to ask how one ought to live, or to reflect on what one holds most valuable. Again, these much more sophisticated, reflective modes of self-awareness require us to construct (consciously or unconsciously) complex self-representations, or models of ourselves. And as many authors have pointed out, there is a narrative structure to such self-representations.

According to Owen Flanagan, there are two basic, and deeply intertwined, aims to these sophisticated self-representations. The first aim is self-understanding. Flanagan asserts that, "when we represent our selves for the sake of gaining self-knowledge, we aim to detect and express truths about the dynamic integrated system of past and present identification, roles, acts, and actional patterns that constitute our 'actual full identity'."[37] The form of this complex self-representation is a narrative; we tell ourselves a story about ourselves, one that might or might not capture the truth of our 'actual full identity'. The second aim of self-representation is for the purpose of social interaction. We project a self-image that could be more or less in line with our own self-understanding and with our actual full identity. Finally, it is widely acknowledge that one's self-representation can affect one's actual full identity.

Now once an individual has developed both autobiographical and contemplative forms of self-awareness, that individual has a complex, robust sense of self. And given this robust sense of self, it would seem reasonable to suppose that there is a real entity, a *self* to which this sense of self answers. Reductionists, of course, deny the validity of this supposition, without denying that human beings typically *do* have a complex, robust sense of self. Rather than positing a real self (or Person) underlying, or standing over and above, the psychophysical complex (including its complex narrative self-representations),

[37] Flanagan (1992), p.195 .

the reductionist takes the self to be a *fictional construct.* The contrast between the realist, anti-reductionist view of the self and the reductionist, non-realist view is well captured by Dennett.

> Two extreme views of the self can be and have been taken. Ask a layman what he thinks a self is, and his unreflecting answer will probably be that a person's self is indeed some kind of *real* thing: a ghostly supervisor who lives inside his head, the thinker of his thoughts, the repository of his memories, the holder of his values, his conscious inner "I." . . . Let's call this the realist picture of the self, the idea of a "proper self." . . . Contrast it, however, with the revisionist view of the self . . . On this view, selves are not things at all but are explanatory fictions. No-body really has a soul-like agency inside them: we just find it useful to imagine the existence of this conscious inner "I" when we try to account for their behavior (and, in our own case, our private stream of consciousness). We might say indeed that the self is rather like the "center of narrative gravity" of a set of biographical events and tendencies; but, as with a center of physical gravity, there's really no such *thing* (with a mass or a shape or a color). Let's call this nonrealist picture of the self, the idea of the "fictive-self."[38]

The idea that the self is a useful fictional construct—a 'center of narrative gravity'—allows the reductionist to understand our sophisticated forms of self-representation and our strong sense of self without committing to the existence of a self. We do not have a self; rather we have a *self-model* (indeed, perhaps more than one such model). This non-realist conception of the self dates back to the Buddhist notion of the *ahaṃkāra,* the 'I-maker'. And as the Buddhists point out, we have an almost irresistible tendency to take this self-model—which is a kind of 'user-illusion', in Dennett's phrase—to be a really existing entity.

Once an individual has developed a full-fledged self-model, this model mediates her self-understanding. Thus, on the non-realist view of the self, a self-presenting mental state presents itself to the cognitive system—and it is therefore available for attention, reflection, memory, action, etc.—but the system represents this occurrence as the state presenting itself to the *self.* With a full self-model, it is phenomenologically *as if* there

were a self or ego behind the stream of experience. In addition, the fictionalist account of the self fits nicely with the idea that our sense of self or identity emerges through process of narrative construction—including perhaps drastic revisions and reinterpretations—and the idea that the self-representations themselves are partly constitutive of our 'actual full identity'. However, one could object that a fictionalist account of the self makes our identity purely a matter of our own creative fancy. What, one might wonder, constrains our narrative constructions if there is no true self to be discovered?

The reductionist answer to this important objection is that, while our ongoing self-representation cannot be constrained by a really existing inner self, it *is* constrained by facts about us, about others, and about the world in general. Even if the self is nothing more than a useful fiction, that doesn't mean I can be the King of England, or a genius, or a well-liked person, simply because I think it to be so. One's actual full identity is *influenced* by, but not completely *determined* by one's own self-representation. The fictive self, like the protagonist in a work of historical fiction, is constrained by the facts. Moreover, even if the self is a useful fiction, that does not entail that the various mental states, dispositions, values, actions, etc., *attributed* to the self are mere fictional constructs. For example, it is common practice to relativize mental states to selves; but even if the reductionist is right that ultimately there are no selves, it does not follow that there are no mental states. So, on the reductionist view, even the fictive self can act as a center of narrative gravity around which we can organize facts (and fictions!) about ourselves. But despite this response, it *is* the case, according to the reductionist, that our common sense, realist picture of the self is ultimately a mistake. There is no inner self to be found in our actual full identity.

[38] Humphrey and Dennett (1989), p.76 .

5.3 Conclusion

According to the reductionist, fundamentally we are not as we seem. Our common sense, substantialist conception of ourselves is mistaken. A person is not a substance, with determinate conditions of identity, existing over and above a suitably organized dynamic system or process of mental and physical events. Terms such as 'person', 'self', and 'I' do not designate independent substances. Instead they are convenient designations of psychophysical complexes. Moreover, there is no inner 'I' that stands behind experience, thinking the thoughts, willing the actions, and watching the perceptual array from within a Cartesian theatre.

However, the central argument of this chapter is not that reductionism is true. Rather, I have shown that a reductionist account of persons is compatible with my view of self-awareness. While perhaps we cannot, as Campbell puts it, 'get beneath' first-person contents in order to characterize mental life from a completely third-person perspective, this fact does not entail that there are persons or selves in any non-reductionist sense. The irreducibility of the first-person can be accounted for by understanding the perspectival nature of consciousness and the way in which mental and physical events are organized within psychophysical systems or processes. Furthermore, one does not need to appeal to persons in order to give an account of the referent of 'I'. Despite our intentions, 'I' might *in fact* refer to a suitably organized series of interrelated mental and physical events, a *skandhasaṃtāna.* Even our deep and sophisticated sense of self can be understood along the fictionalist lines sketched above. The self, on this view, is not a *thing*; it is an explanatory construct, a 'center of narrative gravity'. On the whole, therefore, the general reductionist approach, while certainly revisionist, *is* compatible with the account of self-awareness I have defended.

REFERENCES

Allen, C. and M. Bekoff (1999). Species of Mind. Cambridge, MA, MIT Press.

Anscombe, G. E. M. (2000). The First Person. Self-Knowledge. Q. Cassam. New York, Oxford University Press.

Armstrong, D. M. a. N. M. (1984). Consciousness and Causality. Oxford, Basil Blackwell.

Armstrong, D. M. (1993). A Materialist Theory of the Mind. London, Routledge.

Bermudez, J. L., A. Marcel, et al., Eds. (1995). The Body and the Self. Cambridge, MIT Press.

Bermudez, J. L. (2000). The Paradox of Self-Consciousness. Cambridge, MIT Press.

Block, N., O. Flanagan, et al., Eds. (1999). The Nature of Consciousness. Cambridge, MIT Press.

Brook, A. (1994). Kant and the Mind. Cambridge, Cambridge University Press.

Campbell, J. (1994). Past, Space, and Self. Cambridge, MIT Press.

Cassam, Q. (1992). Reductionism and First-Person Thinking. Reduction, Explanation, and Realism. D. Charles and K. Lennon. Oxford, Oxford University Press.

Casteneda, H. N. (1999). The Phenomeno-logic of the I. Bloomington, Indiana University Press.

Chalmers, D., Ed. (2002). Philosophy of Mind. New York, Oxford University Press.

Collins, S. (1982). Selfless Persons. Cambridge, Cambridge University Press.

Crane, T., Ed. (1992). The Contents of Experience. Cambridge, Cambridge University Press.

Dennett, D. (1991). Consciousness Explained. Harmondsworth, Penguin.

Dreyfus, G. (1997). Recognizing Reality. Albany, State University of New York Press.

Dreyfus, H., Ed. (1982). Husserl: Intentionality and Cognitive Science. Cambrideg, MIT Press.

Duerlinger, J. (1993). "Reductionist and Nonreductionist Theories of Persons in Indian Buddhist Philosophy." Journal of Indian Philosophy **21**: 79-101.

Edelman, G. (1989). The Remembered Present: A Biological Theory of Consciousness. New York, Basic Books.

Eilan, N., R. McCarthy, et al., Eds. (1993). Spatial Representation. Oxford, Basil Blackwell.

Evans, G. (1982). Varieties of Reference. Oxford, Clarendon Press.

Flanagan, O. (1992). Consciousness Reconsidered. Cambridge, MIT Press.

Gallagher, S. and J. Shear, Eds. (1999). Models of the Self. Thorverton, Imprint Academic.

Ganeri, J. (2001). Philosophy in Classical India. London, Routledge.

Gibson, J. J. (1979). The Ecological Approach to Visual Perception. Boston, Haughton Mifflin.

Goodale and Milner (1991). "A Neurological Dissociation Between Perceiving Objects and Grasping Them." Nature **349**(10): 154-156.

Goodale, Jakobson, et al. (1994). "Differences in the visual control of pantomimed and natural grasping movements." Neuropsychologia **32**(10): 1159-1178.

Gupta, B. (1998). The Disinterested Witness. Evanston, Northwestern University Press.

Hattori, M. (1968). Dignāga on Perception. Cambridge, Harvard University Press.

Humphrey, N. C. and D. Dennett (1989). "Speaking for Ourselves." Raritan: A Quarterly Review **9**: 69-98.

Husserl, E. (1931). Ideas: General Introduction to a Pure Phenomenology. London, Aleen and Unwin.

Husserl, E. (1960). Cartesian Meditations. The Hague, Martinus Nijhoff.

James, W. (1950). The Principles of Psychology. New York, Dover.

Johnson, M. (1987). The Body in the Mind: The Bodily Basis of Meaning, Imagination, and Reason. Chicago, University of Chicago Press.

Johnston, M. (1997). Human Concerns without Superlative Selves. Reading Parfit. J. Dancy. Malden, Blackwell: 149-179.

Kant, I. (1964). Critique of Pure Reason. London, Macmillan.

Kapitan, T. (1999). "The Ubiquity of Self-Awareness." Graxer Philosophische Studien.

Kapstein, M. (2001). Reason's Traces. Boston, Wisdom Publications.

Kelly, S. (2000). Grasping at Straws: Motor Intentionality and the Cognitive Science of Skilled Behavior. Heidegger, Coping, and Cognitive Science: Essays in Honor of Hubert Dreyfus. M. Wrathall and J. Malpas. Cambridge, The MIT Press. **2:** 161-177.

Locke, J. (1975). An Essay Concerning Human Understanding. Oxford, Clarendon.

Matilal, B. (1986). Perception. Oxford, Clarendon Press.

McDowell, J. (1994). Mind and World. Cambridge, Harvard University Press.

Merleau-Ponty, M. (1989). The Phenomenology of Perception. London, Routledge.

Mohanty, J. N. (1972). The Concept of Intentionality. St. Louis, Green.

Mohanty, J. N. (1992). Reason and Tradition in Indian Thought. Oxford, Clarendon Press.

Mokṣākaragupta (1985). Bauddha-Tarkabhāśā. Sadanand Bazar, Varanasi, Asha Prakashan.

Moore, G. E. (1903). "The Refutation of Idealism." Mind **12**: 433-53.

Neisser, U., Ed. (1991). The Perceived Self. Cambridge, Cambridge University Press.

Parfit, D. (1987). Reasons and Persons. Oxford, Clarendon Press.

Parfit, D. (1999). "Experiences, Subjects, and Conceptual Schemes." Philosophical Topics **26**(1 & 2): 217-270.

Peacocke, C. (1986). "Analogue Content." Proceedings of the Aristotelian Society **60**: 1-17.

Perrett, R. (2003). "Intentionality and Self-Awareness." Ratio **16**(3).

Perry, J. (1993). The Problem of the Essential Indexical and Other Essays. Oxford, Oxford University Press.

Phillips, S. (2001). Tattvacintāmani of Gangeśa.

Rescher, N. (1996). Process Metaphysics. Albany, SUNY Press.

Rosenthal, D. (1999). A Theory of Consciousness. The Nature of Consciousness. O. F. Ned Block, and Guven Guzeldere. Cambridge, MIT Press.

Rowlands, M. (2001). The Nature of Consciousness. Cambridge, Cambridge University Press.

Sartre, J.-P. (1956). Being and Nothingness. New York, The Philosophical Library.

Sartre, J.-P. (1962). The Transcendence of the Ego. New York, Noonday Press.

Shoemaker, S. (1996). The First-Person and Other Essays. Cambridge, Cambridge University Press.

Siderits, M. (1997). "Buddhist Reductionism." Philosophy East and West **47**: 455-78.

Sommerhoff, G. (2000). Understanding Consciousness. London, SAGE Publications.

Strawson, P. F. (1966). The Bounds of Sense. London, Metheun.

Varela, F., et al. (1991). The Embodied Mind. Cambridge, MIT Press.

Wider, K. (1997). The Bodily Nature of Consciousness. Ithaca, Cornell University Press.

Williams, P. (1998). The Reflexive Nature of Awareness. Richmond, Curzon Press.

Wittgenstein, L. (1958). The Blue and Brown Books. Oxford, Basil Blackwell.

Wittgenstein, L. (1978). Philosophical Investigations. Oxford, Basil Blackwell.

Zahavi, D. (1999). Self-Awareness and Alterity. Evanston, Northwestern University Press.